A New Look at the World

Fractal geometry

A New Look at the World

A New Look at the World

Fractal geometry

Maria Isabel Binimelis Bassa

Everything is mathematical

*To my confidant in fractals Antoni Benseny
and in memory of Juan Femenias*

Contents

CONTENTS

Preface

Forty years ago, no one had even heard of fractals. The concept did exist but the name had not yet been invented. In his book, *The Emperor's New Mind*, Professor Sir Roger Penrose expounds the idea that mathematics is discovered and not invented. Mathematical objects are idealised representations that are often revealed by looking at the order that underlies certain aspects of the world.

The fluffiness of a cloud, the branching of a tree, the irregularity of a bolt of lightning, the ornamental details of a peacock and the calcification of a stalactite are all phenomena that can be described using fractal geometry – in the same way that the Earth can be described by a sphere or the shape of a nautilus by a spiral. However, the Earth is not a perfect sphere; it is not even a perfect ellipsoid. Yet such an approximation gives us a considerable advantage when it comes to making calculations for the prediction of eclipses, and it allows us to predict them to a relatively high degree of accuracy.

When scientists began to question new types of shape that had previously been discounted as irrelevant or ephemeral, they had reach a turning point in the progress of describing and understanding the universe. These new shapes are classified as irregular and sometimes even chaotic. In an attempt to give order to this 'disorder', mathematicians sought and found rules that had until then been ignored – geometric patterns repeated at different scales of observation, both spatial and temporal. It was then, according to Nassim Nicholas Taleb in his book *The Black Swan*, that all the pieces of the puzzles that had been devised since times of antiquity by thinkers such as Plato, Yule and Zipf, were joined together in the hands of Benoît Mandelbrot. According to Taleb, it was Mandelbrot who joined the dots, linked randomness (a randomness that is ultimately a mirage) to geometry and brought the matter to its logical conclusion. To support his theory, he unearthed mathematical works that were unknown at that time.

Shortly after, the range of relationships between this new mathematics and a host of other branches of science such as biology, geology, urban development, economics, technology and even art became clear. Almost simultaneously, the so-called 'Mandelbrot set' emerged from the darkness, a construction that, based on a simple recursive process of squaring, gave rise to an example of surprising complexity and beauty – a structure that does not only live in our minds but has its own inner autonomy, which precedes the moment it was perceived by people for

the first time. Mathematical concepts are timeless entities that reach beyond our existence while, at the same time, constituting a profound truth that goes beyond the material world. From the dawn of humanity, men and women have felt a compelling curiosity for unravelling the laws of the Universe. Fractals represent a new way of seeing the world and inhabit places we are only just beginning to explore.

Chapter 1

The Evolution of Geometry: Mandelbrot versus Euclid

> *"Lines that are parallel meet at infinity!"*
> *Euclid repeatedly, heatedly, urged.*
> *Until he died and so reaching that vicinity*
> *he found that the damned things diverged.*
> Piet Hein, *Grooks VI*

From early times, man has attempted to explain the behaviour and structure of the cosmos. Laws of the Universe have been sought to define the movement of the planets or the shape of the galaxies. Formulae have been sought to predict how an object falls or to describe the flight of birds; and, in the same way, we have studied the anatomy of living beings and the structure of the human mind.

In general, the cosmos is subdivided into two aspects: the macro and the micro. The term macrocosm is used to refer to the collection of everything that forms part of the Universe and exceeds human scale, things in the order of our planet or greater, such as the Solar System, our galaxy or the constellations. In contrast, the microcosmos is the collection of everything on a scale that is less than or equal to ours, such as the organs of our bodies, viruses or molecules. Curiosity has driven human beings to see the invisible or predict random events, and this has led into the abyss of the macrocosm and the finest fissures of the microcosm.

The belief that there is a connection between the macrocosm and the microcosm dates back many, many years. In ancient Greece this connection was based on a belief that the scheme of one thing – the laws and patterns that underlie it – could be found in another, starting with the macrocosm or Universe and working down to the microcosm. After discovering that the ratio 1.6180..., subsequently christened the 'golden ratio', was present in all sorts of natural phenomena, the Greek philosophers attempted to provide a rational explanation for this repetition, hence taking a first step towards a theory of the unification of the macro and the micro. Transcrip-

tions of the *Emerald Tablet* have been found in Arabic texts dated to 650. The original tablet is a Greek text attributed to the mythical Hermes Trismegistus (from which the adjective *hermetic* is derived), which purports to reveal the secret of the 'primordial substance' and its transmutations. The *Emerald Tablet* is a summary of the *Magnum Opus*, the main aim and life's work of an alchemist. Its second precept is as follows "That which is above is from that which is below, and that which is below is from that which is above, working the miracles of one." The precise aim of alchemy is to understand this mysterious relationship between the macrocosm and the microcosm and thus achieve wisdom – and hopefully wealth.

The Greek goal of unifying the micro and the macro has recently been seen in a new light after being viewed from the perspective of a new branch of mathematics. Born from within the extensive family of geometries, this branch is essentially based on two fundamental concepts: self-similarity and continuity. Over time, both concepts have been sketched out and refined, and we shall now turn our attention to these in as great a depth as possible and so study some extremely interesting mathematical constructs in detail further on.

MACROCOSM AND MICROCOSM

The Finnish-born American architect Eero Saarinen clearly and intuitively expressed the requirement for a multidimensional approximation of reality when he said: "Always design a thing by considering its next larger context – a chair in a room, a room in a house, a house in an environment, an environment in a city plan." Saarinen was the source of inspiration for the American architect and designer Charles Eames in the creation of the book (and subsequent documentary) *Powers of Ten*, a succession of images at different scales that begins with our galaxy and progressively approaches an image on a human scale, that of a man tending an urban park, and from then on, it reduces in scale until entering into the microcosm of living organisms and the basic constitution of matter. The centre of each image is the same: galaxies and atoms, macrocosm, all are thus 'connected'. This is an extremely powerful tool for understanding, both theoretically and visually, the idea of a hierarchy of scales. You can find out more about the work on the website http://www.powersof10.com.

Practically all of current science is based on mathematics. It is not in vain that the degree of scientific maturity of a civilisation is often measured according to the level of mathematics that it employs. In this respect, geometry was a basic tool for the technological and scientific evolution of all civilisations, but at the same time it can also be said that technological and scientific requirements lay behind the development of geometry.

We can find a particularly clear example of this symbiosis in the Chinese culture. The Chinese travelled the oceans without fear of getting lost and with the peace of mind that they could return to port thanks to the accuracy of their cartographic projections. Geometry was also extremely useful for the Egyptians when it came to building the pyramids. For his part, the mathematician and astronomer Aristarchus of Samos (310BC–230BC), calculated the distance between the Earth and the Moon with considerable accuracy, and in spite of his rudimentary measurements, the mathematician and astronomer Eratosthenes of Alexandria (276BC–194BC), born in the North African colony of Cyrene, achieved a close approximation of the Earth's diameter. Much later on, with the help of the reflecting telescope, millions of new stars were discovered, and thanks to the technique of stellar parallax it was possible to measure the distances to some of the closest.

According to Herodotus, the historian and geometer who was born in Halicarnassus, geometry was highly advanced in ancient Egypt and the Greeks acknowledged that the Egyptians were responsible for its invention and that they had learnt much from them. All that now remains of Egyptian mathematics are a number of formulae for the calculation of volumes, areas and lengths. We know these were derived for practical purposes, to calculate the dimensions of plots of land, such as for marking fields after the annual Nile flood. This is where the name γεωμετρια or geometry, 'measuring of the earth' comes from (γεω, 'earth', and μετρια, 'measuring'). When Eratosthenes, in a show of his genius, calculated the diameter of the Earth, the word *geometry* had already been invented. However, despite having this other etymological origin, his planet-sized experiment honoured the word.

Geometry studies the properties of space and has traditionally determined rules for the calculation of the lengths, angles and surfaces of the various objects that populate everyday life. The essence of geometry is deeply rooted in our perception of reality. All the information we receive from the world around us, everything we see, hear and touch, is first processed in terms of geometry... and our resulting actions are a consequence of this fact. The plans of our houses and the land that is cultivated are often adapted to a square grid. The path of a falling object, the

ERATOSTHENES AND THE CALCULATION OF EARTH'S RADIUS

Human beings first realised that the Earth was round many years ago, largely as a result of some specific clues. The positions of the stars were considerably different when viewed from locations in the north or the south; when boats disappeared over the horizon, it was always the hull that went first; the Earth's shadow on the moon during a lunar eclipse was round… Eratosthenes, head of the library at Alexandria before Euclid, devised an extremely simple but nonetheless surprising method for calculating the radius of the Earth. From references obtained in a papyrus held in his library, he knew that in Swenet (now called
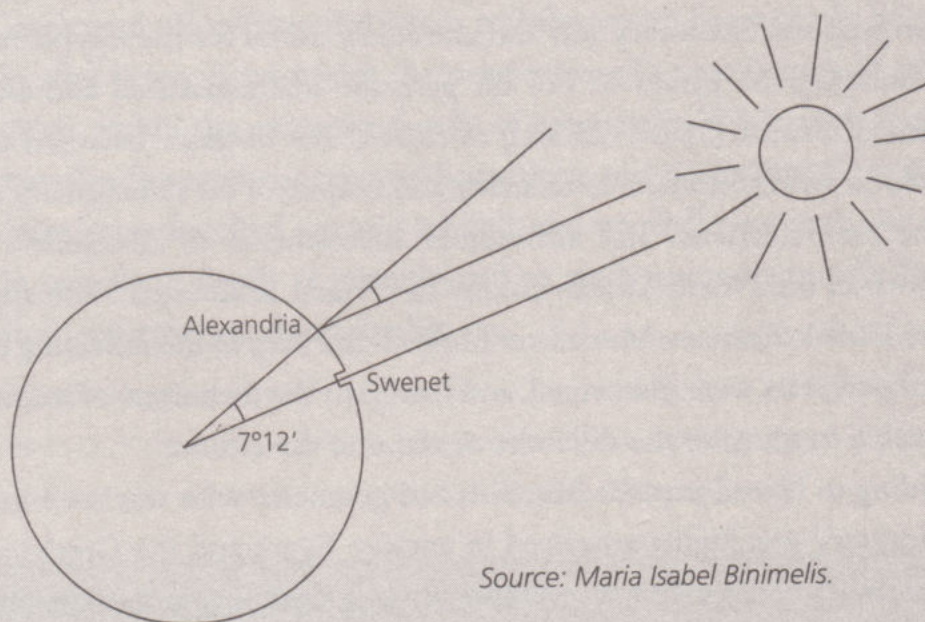
Source: Maria Isabel Binimelis.

Aswan), objects did not have a shadow on the day of the summer solstice and the bottoms of wells were lit up by the sunlight. Assuming that Alexandria lies exactly to the north of Swenet (in fact it is slightly to the north-west) and that the sun is so far away from the Earth that its rays are parallel, on the day of the summer solstice he measured the shadow in Alexandria at noon. By doing so, he showed that Swenet was 7° 12' from Alexandria. Prior to this, he estimated the distance between both cities as 5,000 stades. Armed with these two pieces of information and a little trigonometry, he set about his calculations. If we assume that Eratosthenes took one 'stade' to be a measurement of 185 m, the error was around 6,616 km (approximately 17%). Yet there are also those who claim that he used the Egyptian stade (300 cubits at 52.4 cm each), in which case, the polar circumference would have been 39,614.4 km. When compared with the current estimate of 40,008 km, this is an error of less than 1%.

planetary orbits, the external structure of a Nautilus, the way in which an electrical cable sags... all these shapes are described by formulae and theorems from traditional geometry. Until well into the 19th century, science had not yet opened its eyes to another type of shape.

From what we have seen until now, geometry would appear to be a useful and practical tool. However history also provides us with indications that this is all it was good for. On certain occasions, geometry is used to describe things that are not visible. In his dialogue *Timaeus*, Plato considers the substance from which everything is made. His argument leads him to consider five basic elements: fire, air, earth, water and ether. Based on a series of logical deductions, he associates a geometrical shape with each of them — the tetrahedron for fire; the octahedron for air; the hexahedron or cube for earth; the icosahedron for water; and finally the dodecahedron for ether. These five polyhedra, which were subsequently to be referred to as the Platonic solids, are the only regular ones. (Remember a polyhedron is said to be regular if its faces are congruent regular polygons and the same number of faces meet at each vertex.)
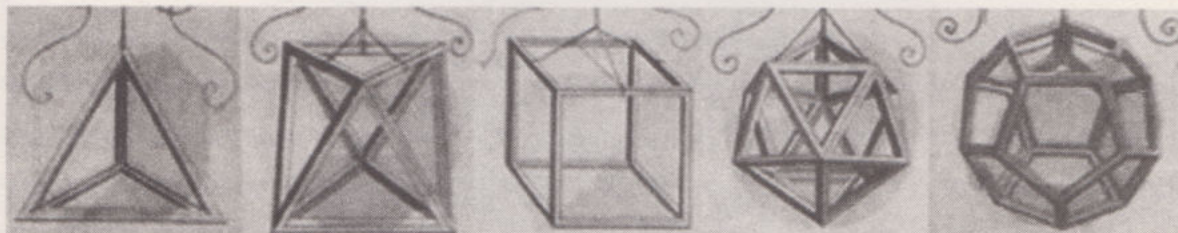
Fragments of drawings of the 'Platonic Solids' designed by Leonardo da Vinci for Luca Pacioli's celebrated De Divina Proportione *(Venice, 1509).*

For his part, the German astronomer Johannes Kepler (1571–1630) spent twenty years in an unsuccessful attempt to fit the Solar System into a harmonious pairing of spheres and regular polyhedra. He proposed situating each of the planets known at the time in six spheres separated by an inscribed polyhedron, thus all the orbits were contained in giant spheres.

The largest is Saturn, separated from Jupiter by a cube; between Jupiter and Mars there is a tetrahedron; between Mars and the Earth there is a dodecahedron; between the Earth and Venus there is an icosahedron; and between Venus and Mercury there is an octahedron. This framework appeared to work reasonably well given the available observations, except for the case of the orbit of Mercury.
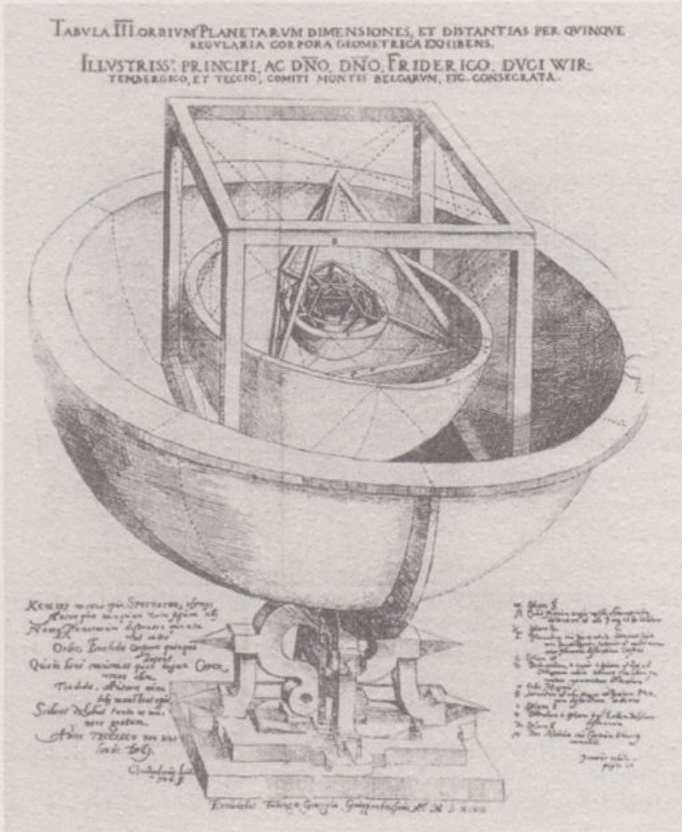
Above, engraving of Johannes Kepler by Frederick Mackenzie. The illustration below shows the Platonic model of the Solar System devised by the German mathematician in his work Mysterium Cosmographicum *(1596).*

Unable to explain the 8 arcminute discrepancy between his theoretical, circular orbit (in his view 'the most perfect trajectory'), and the observation tables of astronomer Tycho Brahe (1546–1601), a somewhat dejected Kepler realised that he would have to abandon the circle. In the end, he made use of the ellipse, which was at the time a strange shape with a mathematical description courtesy of the ancient Greek mathematician Apollonius of Perga, whose work on conic sections wasrescued from the library of Alexandria before its desctruction. And Kepler discovered that it fitted perfectly with the experimental measurements of Brahe.

The terrible deity

Before continuing along the winding route geometry took through history, let us pause briefly to examine the human need to explain our surroundings. In order to do so, it will be helpful to introduce the concept of 'chaos'.

In classical teachings, the cosmos and chaos were considered as opposite elements, and they are often associated with order and disorder respectively. However, as is the case in many creation myths, and in particular the one told by the Greek poet Hesiod, chaos is personified as a deity, the goddess Khaos[1]. In contrast, the cosmos is not identified with a divine figure. Chaos is derived from the root *ghen*, which comes from the Proto-Indo-European language and means 'chasm' or 'gaping', possibly because in the context of the time when the term was created, chaos was understood to fill the space between the heavens and the earth. Only later on, as a result of linguistic evolution, did it come to mean disorder.

From the earliest days of humankind there have been phenomena that we do not understand, that arouse surprise or fear, and these are often jointly referred to as the 'world of chaos'. As Hesoid explains in his *Theogony*, written around 2,700 years ago, the gods arose precisely to impose order within chaos and thus put an end to human insecurity. The history of science also involves the quest for order within chaos, and mathematics has proved an indispensable tool in this search.

Just like the gods, men have attempted to give order to chaos. In his time, Plato had already stated that nature was mathematical, and over the centuries, the majority of scientists held a similar belief. The search for mathematical explanations for those natural manifestations that appear to subscribe to a certain order is an ancient one. However, what do we do with things that do not conform to rules? According

---

1 In contrast to Hesiod, the Orphic tradition considered Khaos to be a descendent of Chronos and Ananke.

to the popular science writers Ian Stewart and Martin Golubitsky, chaos provides new rules of regularity based on an apparent disorder. However, as much as we focus on the definitions, literally speaking, chaos should not have any type of rules. Stewart and Golubitsky are in fact referring to a certain special type of chaos: The mathematical concept that arose from the framework of dynamical systems theory. The idea that chaos, understood as the radical absence of rules, does not exist, is part of our primitive belief that there is a general theory that explains everything, be it theogony, alchemy… or science.

IN THE BEGINNING THERE WAS KHAOS
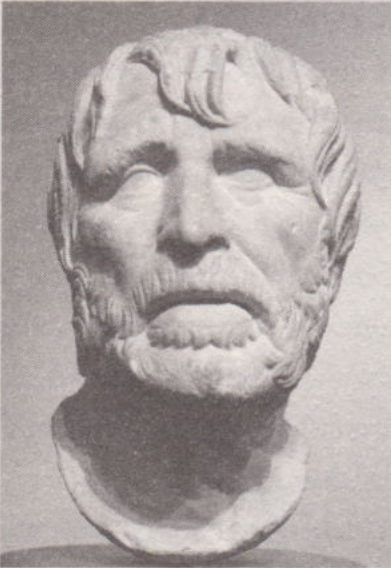
For many years it was thought that this bust was of the Roman philosopher Seneca. However today it is believed to represent Hesiod.

The *Theogony* written by the Greek author Hesiod (8th century BC) begins by recounting the birth of the Universe from Khaos, a distant, infinite abyss where elements moved without direction. Two indefinite sister entities coexist in the heart of this primordial abyss, Erebos (Darkness) and Nyx (Night). Upon being separated both from each other and from Khaos, their mother, they give birth to their opposites and complements: Aether (Light) and Hemera (Day). Day and Night join together to form Time. Then, but not descended from Khaos, comes: Gaia, the element of stability, the universal mother who confronts Khaos and gives birth to all that exists: Eros, love, first creator of life and Tartaros, the Underworld. For her part, Gaia gives birth, without the intervention of a male partner, to all that is still absent from the Universe: Uranus (the Sky) to cover her and surround her completely. With the Sky and Earth pair already formed, she organises the world into a symmetrical and balanced cosmos.

In keeping with the literal sense of the word, the *Theogony* should have been the tale of the birth of the gods. However Hesiod begins with a cosmogony, that is to say, with the birth of the 'ordered universe'. The birth of the gods, as he explained it, came later.

A snapshot of complexity, urban development and linguistics

Most scientists would agree with Paul Dirac, British Nobel laureate in physics, who said that, "Physical laws should have mathematical beauty". The desire to explain the entire Universe with a logical, simple and beautiful theory has always been a fundamental one. Plato's theory was one of the first known to associate matter (the five elements) and geometry (the five polyhedra). Many years later, they would be succeeded by the scientific and mathematical theories of figures such as Hendrik Lorentz, Henri Poincaré, Hermann Minkowski and Albert Einstein, who proposed general ideas that added space and time to matter and geometry.

However, if the scientist's requirement is to find a general law that is both simple and beautiful, is it possible to claim, in some respects, that humanity naturally tends towards simplicity. The circle is by far the most common geometric shape, closely followed by the square, the triangle, the pentagon and the other regular polygons. Then there are conics (the ellipse, the hyperbola and the parabola), spirals and other noteworthy curves. These shapes are believed to be the 'purest' ones. They have been almost idolised throughout time and invested with great symbolic meaning, defining the environment of all human settlements. They are surely archetypical figures in the sense used by the psychologist Carl Jung, and have been used profusely as shapes, either sacred or profane, throughout the ages. In architecture, for example, the circle appears at Stonehenge and in Roman amphitheatres and basilicas; the square and the triangle in pyramids and ziggurats; the pentagon in numerous decorative patterns; the list goes on.

However, there is also another type of configuration. Ron Eglash, an anthropologist at the Rensselaer Polytechnic Institute in New York, has investigated architecture, sculpture, textile patterns, games and other cultural shapes from Africa. Eglash has observed that the patterns that occur in different cultures can be characterised by specific designs. In Europe and America, we often see cities based on a grid of straight line streets with right-angle corners. In contrast to this, traditional African settlements tend to make use of mixed structures such as rectangular walls enclosing increasingly small rectangles, and have streets with wide avenues branching out into narrow paths with intricate and geometric repetition. These indigenous structures are not limited to architecture, their recursive patterns are echoed in a wide variety of designs. Eglash also discovered that many towns throughout Africa were designed to reflect the social hierarchy. The status of residents determined where their house was located within the town.
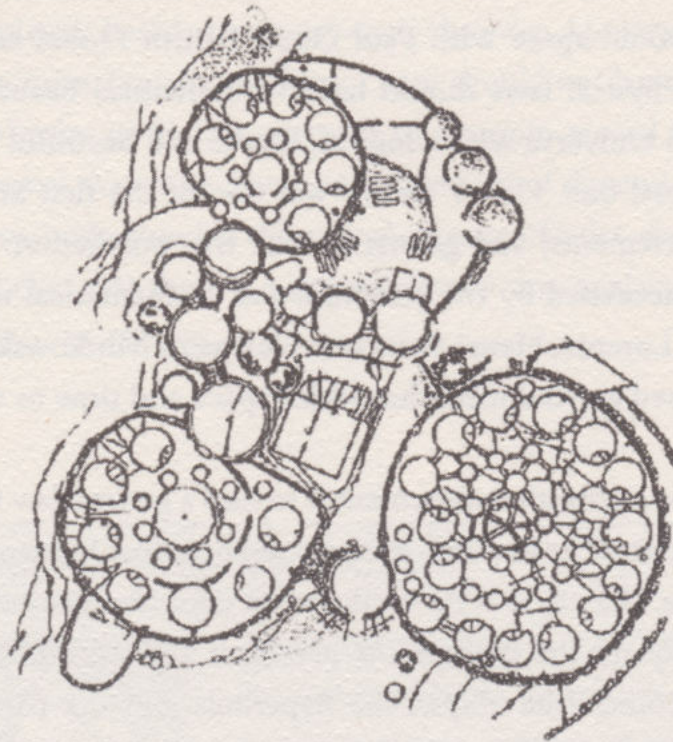
The anthropologist Ron Eglash discovered various types of fractal settlements, some based on rectangles, others on circles, as in the top picture. The middle picture shows a Cameroonian settlement with self-similar morphology, and the bottom one shows a morphological study of various products from native Cameroonian culture.

In spite of some evidence of certain fundamental contrasts, such as those discovered by Eglash, it is clear that the medieval nuclei of most European cities also exhibit highly irregular forms. It is believed that such urban models have historically been developed like a biological organism, that is to say in order to optimise a series of vital activities.

This hypothesis is based on complexity theory and shows how these societies have taken advantage of the non-linear aspects of their organisation in order to maintain a series of relationships with their environment. Certain dynamic ecological systems have been imitated without realising it. Colonnades, arcades, rows of narrow buildings with crossed paths... all these architectural elements are fashioned in the image of a permeable membrane with cavities to allow for exchanges. The greater the segmentation (the more cavities), the greater the transmission of information.
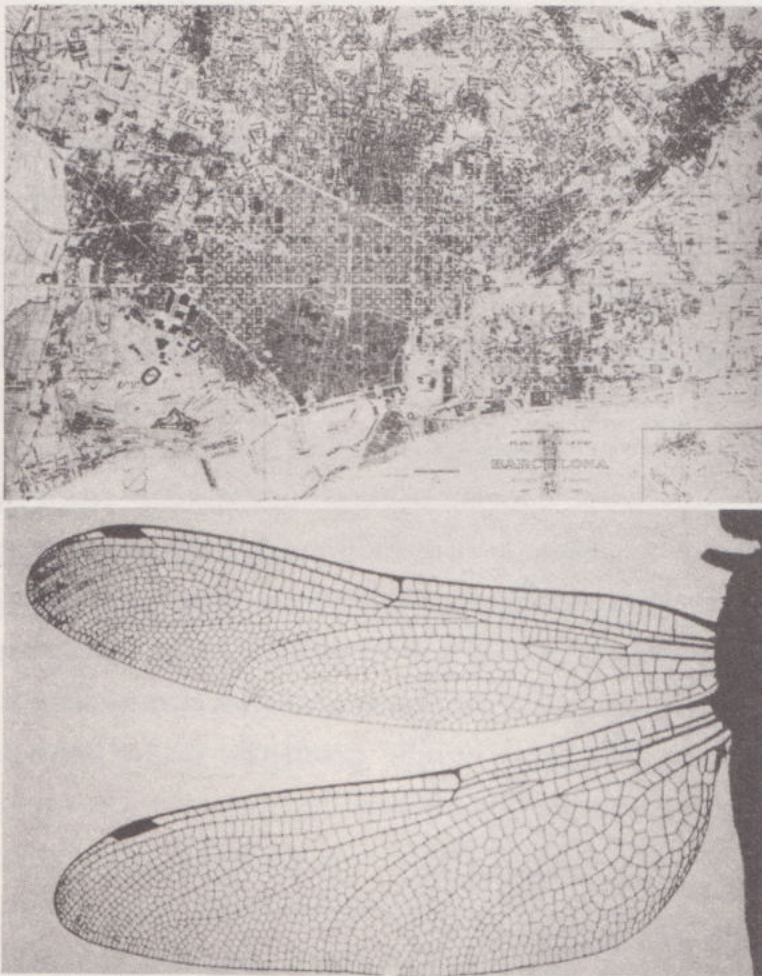
The layout of a traditional Western city, such as Barcelona, exhibits a number of similarities to the networks found in certain animal, such as the wing of an insect.

SPANISH CITIES: A MELTING POT OF INFLUENCES

Just like living organisms, cities can be thought of as tissues that evolve over the course of time. The urban space, understood as the transformation of rural space, and thus of the natural, entails the creation of two different typologies: streets (empty space) and plots (occupied space). The morphology and development of these tissues are products of urban development activities heavily influenced by culture. In Spain, for example, cities follow a model that corresponds to four periods of development:

- The classic Mediterranean city derived from the Graeco-Roman tradition.
- Medieval Muslim and Christian cities, followed by the transformations of the Renaissance, the Baroque and the Enlightenment.
- The preindustrial modernist city, followed by the industrial.
- The postindustrial and suburban city.

In these urban environments, the Eastern and Western influences are representative of a synthesis that converts their old towns into a unique space quite different from the remainder of European metropolises. They are extremely compact and structured dendritic networks (meaning they have many branches) that spread out from typical focal points such as fortresses, mosques, churches and convents. The complete city is clearly bounded by a structural element, such as a the wall.

As the original, primitive activities of certain cities has changed to suit different purposes, it is possible to see a metamorphosis in the initial town layout. Renovation plans do away with what was originally needed. For example, whereas at one time exchanges between small groups were most important, now speed and the transfer of large numbers of people is most needed. Large-scale planning based on 'circular' beltways, ring roads and rectilinear grid-iron layouts are put into action to prevent the city from failing. Are 'simple' geometric shapes the best ones for this task? Is is possible to plan something better?

Cities, more than just maps

Thus the shapes of cities are the result of a long, repeating construction process that interacts with the physical site, in which many people have intervened and taken decisions independently that have culminated in a modern city – an entity that can

no longer be understood using Euclidean geometry. A multidimensional approximation is required, that is to say, one that entails observation from different distances that preserves the relationship between all points of view.

Observing the city using different scales, we can discover that the shapes observed at one scale are similar to those seen in detail in other ones. In this way, we can conclude that there are basic similarities between the branching of trees and urban road networks. Both are produced by several iterations of divisions and this suggests that natural growth and urban development are produced through a combination of iteration and chance.

The way the modern Western city expands is an expression of bourgeoisie rationality and order, a mind-set that reflects a boundless Euclidean logic that dominates the entire land. We always try to apply Euclidean shapes (circumferences, squares, cubes, etc.) to the real world, despite the fact that they are confined to the realm of the purest mathematical abstraction and, as such, our minds. In the end however, reality resists such orderings and simplification, and ends up reasserting the complexity of its own nature. The result is an irregular geometry based on the relationship between the objects and their many parts on different scales. In order to get to grips with this irregular reality, we need to find an alternative geometric model that is based on these relationships, without resorting to ideal geometric shapes.
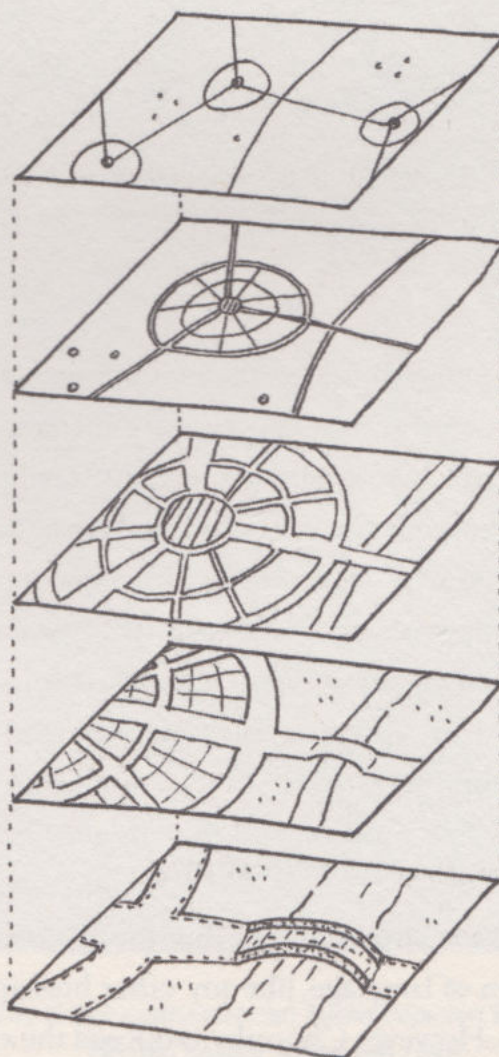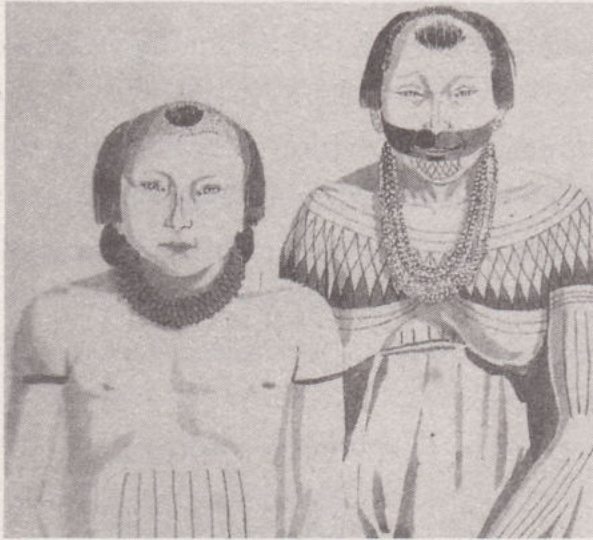

Successive approximations of an urban structure, which shows the shapes of the larger and smaller scales are self-similar. (Source: Laura Elisabeth Violant.)

Geometry and linguistics: is geometric thinking innate?

In spite of what we have discussed previously, when land is in demand and specific boundaries must be marked out, we normally make use of a rectangle or quadrilateral, as is the case when dividing the countryside into fields.

It appears that we have always made use of the quadrilateral and in fact it is one of the most commonly adopted shapes, together with the circle, the spiral and the cross. In this respect, some researchers have sought proof that geometric concepts are innate to human beings and do not require language or culture in order to appear. This has been researched by carrying out comparative tests on children and adults from the Munduruku tribe in the Amazon, which remained isolated from our civilisation for more than four centuries after the arrival of the European conquerors. Their primitive knowledge of geometry shows that our species has some geometric intuitions that are independent of culture and education, rooted in an experience

Members of the Munduruku tribe as depicted by the French artist and photographer Hércules Florence in 1828.

that precedes a knowledge of maps or graphical symbols, and even a language of geometric terms. This revelation constituted a revolution in neurology, anthropology, psychology and hermeneutics, all disciplines that have still been unable to determine if language is the driver for such common knowledge.

Modern linguistics have revealed universal semantic features that are common to all languages, similar sounds that make up the words of all spoken languages. Is this the key to determining if, in the same way as with language, there are also universal geometric or arithmetic features shared by human beings? Does knowledge precede the subject? Is it inherited, is it innate? Are these possibly genetic, like the linguistic universals proposed by the eminent linguist and philosopher Noam Chomsky? In 1957, at just 29 years old, Chomsky revolutionised the field of theoretical linguistics with the publication of *Syntactic Structures*. Until then, it was believed that the acquisition of language, like any other human skill, was a product of learning and association. However Chomsky proposed the existence of an innate device in our brain, the

'language organ', which allows us to learn and use language almost instinctively. Moreover, he proved that the general principles of grammar are universal to all human cultures and proposed the existence of a universal grammar.

The prominence of Euclid

Around 323BC, at the time of the death of Alexander the Great, the fame of Greek science had reached all the countries it had conquered. It is not strange then, that Ptolemy I, pharaoh of Egypt, came to Athens in search of experts in sciences at a time when Alexander's greatest city, Alexandria in Egypt, was emerging as a great centre of culture. Euclid was nominated to direct the school of mathematics.

The first philosopher to mention Euclid was Proclus in 800, who dates the birth of the Greek mathematician to around 300BC. However, while this historical fact is uncertain, what is unquestionable is that it was Euclid who gave order to all mathematical knowledge up to that time, perfecting some work and providing irrefutable proofs of things his predecessors had been unable to state with sufficient rigour. His work is a compilation and, above all, a systematisation of the geometry. Prior to Euclid, mathematics was a series of disjointed fields. His work joined them together to form a collection of inter-related systems.

We know that Euclid wrote twelve works, only five of which have survived: the *Elements*, *Data*, *Divisions of Figures*, *Phaenomena* and *Optics*. The *Elements* was a standard textbook at all universities and centres of mathematical study for more than two thousand

The Greek mathematician Euclid depicted by the Flemish painter Justus van Gent.

years[2]. It is thought that some one thousand five hundred editions have been published in languages including Greek, Arabic and Latin. Until the second half of the 20th century it was one of the best-selling books of all time.

The *Elements* is one of the most beautiful, extensive and oldest scientific works that has been handed down to our time. It comprises 13 volumes: six on plane geometry, three on arithmetic, one on measurements and another three on spacial geometry. Its objective is to set out the foundations of mathematical knowledge without practical application and the work achieved a degree of perfection that meant it was necessary to wait until the end of the 19th century to surpass it[3]. His theorems were statements of 'truths' about the world in which we live, and nobody was able to imagine a different geometry.

In order to try to investigate the reason Euclid dedicated so much effort to such an exhaustive and rigorous compilation, it is necessary to go back to the Pythagoreans. Upon studying the length of the diagonal of a unit square, they discovered that the resulting number, $\sqrt{2}$, was irrational. That is to say it cannot be expressed as a fraction of two whole numbers. In the language of the time, the diagonal was said to be incommensurable with respect to the sides. This fact, which no longer surprises us, was considered by some Greeks, including Pythagoras, as such a disaster for mathematics that it even managed to shake the foundations of their cosmology. Euclid, who was familiar with the knowledge of the Pythagorean school and especially with this greatest of crises, wanted to place geometry on solid principles that, together with an unbreakable logic, would establish a series of enduring valid results.

However, a small logical problem arose for him to solve. All proofs start from one or more hypotheses to obtain a result known as a theory. The truth of a theory depends on the validity of the argument with which it has been derived and the truth of the hypotheses (this had also been studied by Aristotle in his *Logic*). In order to be able to determine the truth of the hypotheses, it is necessary

---

2 However with modern standards of rigour, Bertrand Russell claimed that the fourth proposition of the *Elements* was a "meaningless step" and also declared that it was scandalous that the work was still used as a textbook in his time.

3 In 1899, the German mathematician David Hilbert wrote *The Foundations of Geometry*, a work which, based on 21 axioms or postulates, proved theorems from elementary metric geometry and corrected certain defects in Euclid's work.

to consider each as a theory of another argument the hypotheses of which must also be checked. This results in an apparently endless process in which all hypotheses become theories to be proved.

Euclid realised that not all mathematical statements can be proved and that some must be accepted as basic assumptions. In the *Elements* he establishes an example of what is known as the axiomatic method, an important turning point in the history of mathematics. Euclid works with four types of hypotheses: definitions (a total of 23 which can be used to recall the meaning of certain terms), postulates (a total of 5) and axioms (or common notions[4], also 5).

The rationale behind the aximatic system, also referred to as deductive, has been the guiding rule of all modern mathematics. However some of the propositions and theorems declared appear to be flawed when subjected to rigorous analysis. For example, Euclid sets out a principle that has since been shown to be false: *Only* one straight line passes through two points. Nevertheless is took more that 2,000 years to expose these errors at start of the 19th century[5].

One of the contents of the *Elements* that has been the source of the most studies and controversy is the fifth or parallel postulate: "Given a point that does not lie on a straight line, only one parallel line can pass through it[6]." This postulate had the appearance of a theorem – it seemed to be something that was provable. Scholars and commentators on the *Elements* were already aware of its less than intuitive nature, and it was used sparingly by Euclid himself, as if he wished to avoid doing so – he deduced some thirty theorems before relying on it. This leads us to think that Euclid had tried to prove it and, convinced he was unable to do so, simply added it to the others.

---

4 According to Proclus, the expressions 'common notion' and 'axiom' are synonymous for Aristotle and other logicians although axioms *(axiómata)* are never mentioned in the *Elements* and Aristotle prefers to talk of principles *(archaí)* or common premises *(tà koná)* or primordials *(tà prôtal)*. The form of the postulates and common notions appears to geometric, although, the latter are supposedly common to all mathematics. Common notions express fundamental properties of mathematical objects, whereas postulates state the geometric operations that can be carried out. By postulate, we understand what is conceived, what is expressly stated to be possible. Today an axiom is not an evident truth but a logical expression (supposedly clear) used in a deduction. It appears that 'postulate' has remained as an archaic synonym of 'axiom'.

5 Another error committed by Euclid was to omit at least two more postulates: two circumferences separated by less than twice their radius intersect at two points. Two triangles with two equal sides and the angle formed by these likewise, are equal.

6 The original equivalent statement is as follows: "If a straight line intersects another two on the same plane in such a way that the inside angles of one side add up to less than two right angles, these two straight lines intersect at a point located on the same side of the angles in question." .

This aroused a desire to correct this 'defect' among subsequent generations of mathematicians who have tirelessly sought to prove the postulate. Their efforts continued for twenty centuries. All those who believed they had proved it had instead merely introduced another unproven postulate[7]. These unsuccessful attempts meant that its proof became the fourth most famous problem in Greek mathematics, together with squaring the circle, angle trisection and duplicating the square. However it was not until the 19th century that Carl Friedrich Gauss and Nicolai Lobachevski finally proved that it could not be proven at all. This extraordinary discovery shattered the certainty that Euclid's geometry was the one and only possible and paved the way for so-called non-Euclidean geometries, which we shall discuss further on.

The birth of the infinite straight line

During the Renaissance, the new requirements of artistic and technical representation led certain scientists and artists to investigate new geometric techniques that would provide them with a better way of representing reality. Figures such as Filippo Brunelleschi, Leonardo Da Vinci and Luca Paccioli emerged, to name just a few of the best known, whose ambition was to create a sense of depth in their two dimensional works. Their efforts gave rise to the practical foundations of what would later be referred to as 'perspective'.[8]

As has been previously mentioned, the mathematics of Euclidean space is one of the keys to modern thought in areas as diverse as the natural and human sciences, and art. For Euclid, mathematical space was conceived as an empty and absolute space upon which not only pictorial and artistic representations but also reality itself was constructed. This is the space of perspective, which would be impossible without the wonder that is Euclid's mathematics, or rather, without the mathematics of linear space.

---

7 A false proof, attributed to Thales of Miletus, is based on the hypothesis that there is a rectangle with four right angles. However it is impossible to show the existence of rectangles without validating the parallel postulate.

8 The Latin etymology of perspective is derived from the verb *perspicere*, 'to see clearly', which is equivalent to the Greek word *optiké*, 'optic'. Thus perspective originally referred to the study of phenomena related to vision, and this was its meaning in the classical and medieval world. What is understood as perspective from the Renaissance onwards was the subject of another discipline in ancient history: *scaenographia*, which dealt with both architectural drawings and theatre scenery.

Raphael's fresco *The School of Athens* is a famous example of the application of perspective, which depicts all the major Greek sages. In it we can see a large number of philosophers from the classical tradition gathered together in the same space and surrounded by a great architectural frame. This work is, among other things, a trick of time, which has somehow been frozen for all the philosophers in a pictorial eternity. And... there's the old man Euclid himself! Raphael placed him on the right of the fresco, crouched down, surrounded by pupils to which he is lecturing while using a compass to trace arcs on a small slate. Pythagoras also appears in the fresco, sitting in the opposite corner, writing on a tablet. Pythagoras and Euclid are positioned in the bottom corners of the work and are the point of departure for two imaginary straight lines, which tend towards the centre of the composition, precisely where Plato and Aristotle are located, and onward to be lost on the horizon as they disappear into infinity.

The School of Athens: *alongside Euclid and Pythagoras, Raphael's fresco also includes the following figures: Zeno of Citum, Epicurus, Anaximander, Averroes, Alexander the Great, Xenophon, Hypatia, Parmenides, Socrates, Diogenes of Sinope, Plotinus, Archimedes, Zoroaster, Claudius Ptolemy, Protogenes and Raphael himself, represented by Apelles.*

A step towards the infinite: projective geometry

The formal foundations and basic principles of what is known as projective geometry were provided by Gérard Desargues (1591-1661). The French mathematician observed – as any one of us can – that when a circle is viewed in perspective, it takes the shape of an ellipse, and that the shadow projected onto a wall by a circular object will appear circular, elliptical, parabolic or a branch of a hyperbola, depending on the inclination of the object. (The four types of curves just mentioned, the circle, the ellipse, the parabola and the hyperbola, are known as conic sections.) However, the last of these is the same as saying that the projection of an object (in our example, its shadow) transforms one shape into another[9].

Based on these ideas, Desargues was able to simplify geometrical objects by introducing two new concepts: the point at infinity, or the ideal point, and the line at infinity, or the ideal line. There are ideal points on the plane, each associated with a specific direction; all these make up the straight line at infinity. Similarly, in space there is an infinite number of ideal straight lines, as many as there are directions, and all these make up the plane at infinity. According to his model, a pair of parallel straight lines meet at the point at infinity, an ideal point given by the direction of the gradient of the straight line. That is to say, a point at infinity can be associated with each gradient.

In the same way, two parallel planes intersect at a straight line at infinity, that is to say at an ideal line. As such we can say that two straight lines on the same plane (coplanar) will always share a common point (real or ideal), and that two planes in space will always share a common line (real or ideal). The parabola is an ellipse with an ideal point and the hyperbola is also an ellipse, this time with two ideal points. This is the basis of the principle of duality relating points and lines for each correct theorem. If in such a theorem, 'points' are changed for 'planes' and 'pass through' for 'intersect', another valid theorem is obtained. Thanks to the new model, the theorem "only one straight line passes through two points that do not coincide" has its double "two straight lines that do not coincide only intersect on one plane".

---

9 In technical terms, the term 'projection' refers to the geometric transformation that leaves the harmonic quadrilaterals unchanged, or rather the proportion of the segments determined by four points, A, B, C and D such that $AB/CB=DA/DC$).

Hyperbola

Parabola

Infinite line
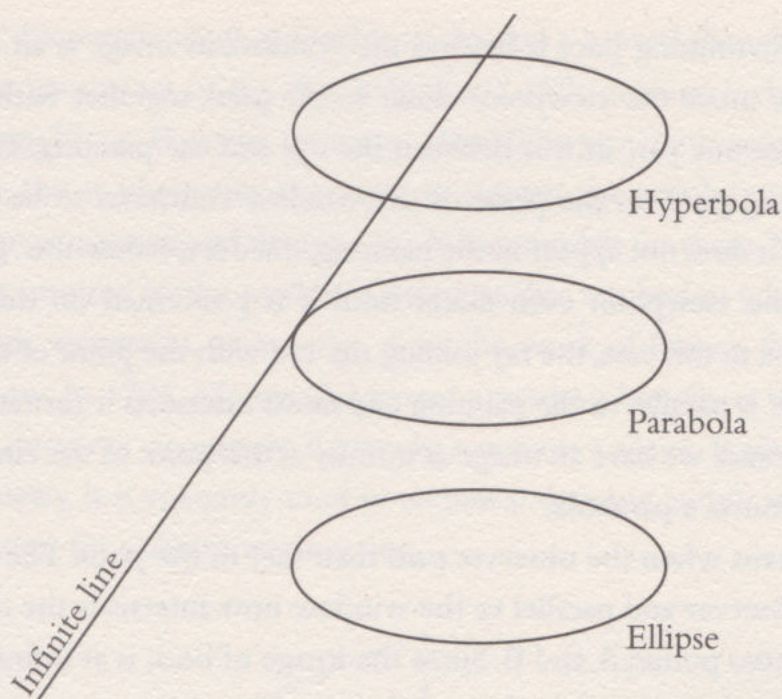
Ellipse

Up to this point, that fact had never been stated in such a way. It was not difficult to check that an ellipse (closed curve) could be the image of a circle in perspective. By means of example, Durero had constructed the flat sections of a straight cone, point by point, for all possible section positions. However even so, in one of his drawings, it can be observed that what was theoretically an ellipse is really drawn in an ovoid shape, as if he did not believe his eyes and expected, to the contrary, that upon approaching the vertices, the cone of the curve was more pointed, instead of a genuine ellipse. In contrast, it seemed inconceivable that open curves with infinite branches, such as the parabola, or those with breaks in their continuity, such as the hyperbola, which has two separate branches, could be assimilated into the circle.

Painting a canvas of a swimming pool

To better understand the way in which the various conic sections appear as projections of a circle in the context of perspective, let us imagine a situation in which the painter wishes to capture a circular swimming pool in a certain part of their painting. They observe the swimming pool through a fictitious window on which they wish to draw it (the image projected onto the window will be the one that will be reflected by the painter in their work). It is possible to make a range of conics appear just by varying the angle at which the window is tilted, however instead let us keep it perpendicular to the ground and modify the positions of the viewer and the window with respect to the swimming pool.

When the swimming pool is beyond the window, its image is an ellipse. However let us now move the viewpoint closer to the pool, together with the window such that at least one part of it is between the eye and the painting. The projection of the swimming pool on the plane of the window continues to be an ellipse, although part of it does not appear in the painting, since it is below the "ground level". Let us move the viewpoint even closer until it is positioned on the edge of the swimming pool; in this case, the ray joining the eye with the point of the swimming pool at our feet is parallel to the painting and never intersects it (or rather it does so at infinity). Because we have an image at infinity at this point of the circle, the swimming pool becomes a parabola.

What happens when the observer puts their feet in the pool? The plane passing through the observer and parallel to the window now intersects the circumference of the pool at two points, A and B. Since the image of both is at infinity, the part of the circular pool that extends in front of the observer produces an open curve with two asymptotes parallel to the straight lines that join the eye with the points A and B. With respect to the arc of the pool behind the observer, although they are unable to see this, it is possible to project it onto the window to give another equal but symmetric curve which is merely one branch of a hyperbola.
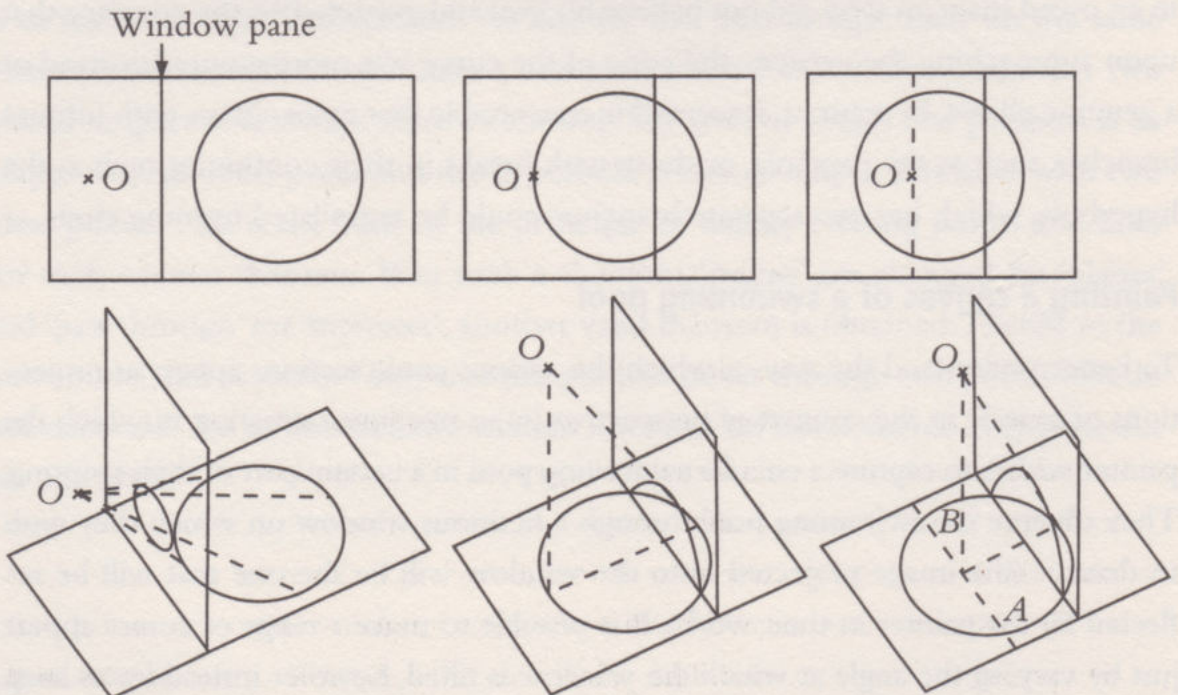

Diagram of the representation of a circular pool depending on the position of the observer. (Source: Maria Isabel Binimelis.)

Desargues' discoveries made it possible to develop a general theory of projections studied by the geometers of the first half of the 19th century, including Gaspard Monge and Jean-Victor Poncelet (to name a few). The projective geometry that was built based on this mathematics also came to suggest the possibility of developing non-Euclidean geometries and devising Euclidean models for them. The first works on perspective inspired by the work of Desargues that resulted in what is known as cavalier, or high viewpoint, perspective were the work of French theoreticians in the 17th century. In 1794, Monge discovered a method known as the orthogonal projection to represent geometric figures in space on a plane. Referred to as descriptive geometry, it is currently used in technical drawing and in its time constituted a revolution for military engineering.
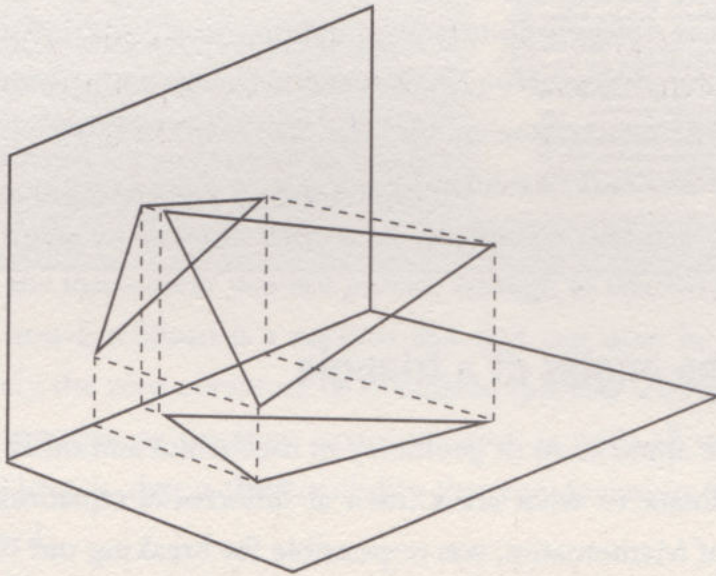
*Orthogonal projection. (Source: Laura Elisabeth Violant.)*

Its use in architecture did not come until later. Cavalier and axonometric perspective (the representation of three dimensions using the projection onto three axes traced on the plane of the drawing) are from the 19th century.

Desargues' contribution can be summarised as follows: whereas Renaissance parallel perspectives considered visual rays to be parallel, with Desargues they meet at infinity. Put another way, cavalier perspective is the same as a central perspective in which the eye of the drawer is 'sent to infinity'. For the Russian Suprematist painter El Lissitzky this appeared to dispense with subjective painting, given that the viewpoint no longer existed. The painter assumed the role of the creator as if looking from infinity.

THE INTRODUCTION OF COORDINATES

The work of René Descartes and Pierre de Fermat, key figures in what is referred to as analytic geometry, ushered in the discipline's so-called Modern Age. Their great contribution was the introduction of coordinate axes, a device that made it possible to give a numerical expression to the points of a geometric shape which could then be studied and treated using algebraic methods. In this manner, a geometric problem could be reduced to an algebraic one, which once worked out, would give the solution to the initial problem.
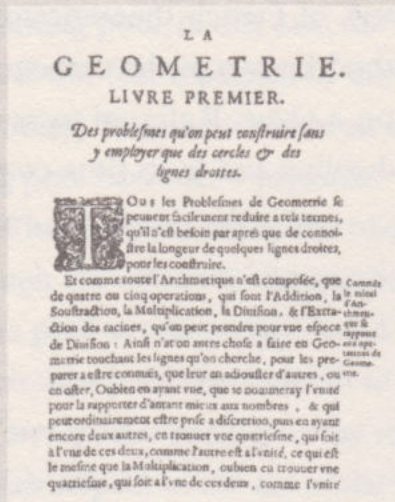
A year before the publication of Descartes' La Géométrie *Fermat had already written: "Whenever in a final equation two unknown quantities are found we have a locus, the extremity of one of these describing a straight line or curved".*

Summing the angles of a triangle

For a while, the application of geometry in mechanics and other branches of physics was subordinate to what are known as differential equations[10]. Gauss, the so-called Prince of Mathematics, was responsible for breaking out of this paradigm by laying the foundations for geometry that studied differentiable variables, or rather differential geometry. Gauss studied curves and surfaces in space and established the fundamental notion of the curvature of a surface. On a circumference, for example, the curve decreases with the size of the radius. Extrapolating from this, the curvature of a straight line would be zero, as was to be expected.

For a surface, however, the situation is more complex. At any rate, we can say, in general and imprecise terms, that given two points on a surface, the shortest curve between them that does not leave the surface (referred to as a geodesic curve) is the curve with the least curvature (i.e. the straightest). This statement is an extension of one that states that on the Euclidean plane a straight line is the shortest line

---

10 In general terms, they are algebraic equations involving derivatives and define the behaviour of flows or fields of velocities.

between two points[11]. Under this interpretation, Euclidean space becomes a specific case of a space with a curvature of zero.

Based on these considerations, Gauss showed that there are surfaces for which the sum of the angles of the triangles formed by geodesic lines is greater than 180°, and that there are others in which it is less. Given that there is a provable equivalence between Euclid's fifth postulate and the fact that the interior angles of a triangle add up to exactly 180°, Gauss' discovery represented a contradiction of the fifth postulate. Towards 1824, and as described in his notes, Gauss had reached the conclusion that it was impossible to prove the fifth postulate based on the others for the simple reason that it was independent of them. Moreover, it was possible to construct a perfectly logical geometry that negated the postulate without contradicting any of the other four. However, although at this point he was already the most prestigious mathematician in Europe, Gauss considered that the mind-set of the age was not prepared for a result of such magnitude and never published these notes.

Some people claim that Gauss was the first person to consider the possibility that the geometry of the Universe was 'non-Euclidean'. It is said that he climbed the summits of three mountains with a theodolite to measure the angles they formed although the instrument was not precise enough to resolve the matter. That physical space is non-Euclidean is a curious fact that can now be shown experimentally. Effectively, the zero curvature of Euclidean spaces is a limit case that separates spaces with positive or negative curvature, and, as with all measurements, is subject to error. There is always the possibility that the deviation from zero is too small to be detected.

Shortly after Gauss, another person came to the same conclusion, however in this instance they were bold enough to publish their findings. The result was a work of geometry that was derived from replacing Euclid's fifth postulate with another: "At least two parallel lines pass through a point lying outside a straight line." In fact, there were two mathematicians who arrived at the same result independently – the Russian Nikolai Lobachevski and the Hungarian János Bolyai, son of Farkas Bolyai, a friend of Gauss. Gauss recognised what he had already developed in his work and was so interested in Lobachevski's exposition that, at 62 years of age, he began to study Russian so as to be able to read the work, which he managed within a few months, genius as he obviously was.

---

11 Fermat's proof that the principle that governs the propagation of light corresponds to the minimum time and not the shortest distance is another experimental proof of physical activity that does not depend on the equality of angles but only the proportionality of the sine.

Bolyai and Lobachevski did not attempt to prove Euclid's parallel postulate starting with the others, as Johann Lambert had done before them, in contrast to many of their predecessors. Instead, they noticed that the parallel postulate must logically be independent of the others. They argued that independence must mean much more. Removing the parallel postulate and including a new, alternative postulate on parallel straight lines (including one stating the opposite) would give another system

ABSOLUTE GEOMETRY

Absolute geometry is the term that describes the collection of geometric results deduced from Euclid's first four postulates. These results are described as absolute since they must hold for both Euclidean and non-Euclidean geometry. As we have already seen, the only difference between the former and the latter is the fifth postulate that deals with parallelism.

The so-called Saccheri quadrilateral, named after the Italian Giovanni Saccheri, and the Lambert quadrilateral, named after the German Johann Lambert, are highly significant as both were used unsuccessfully in attempts to prove the fifth postulate. Saccheri worked with the idea of finding a contradiction when the postulate was denied, that is to say, he tried to prove the fifth postulate by showing its denial would lead to a contradiction. However, he made the mistake of considering certain results to be contradictory or false for the simple fact that they were contrary to normal Euclidean intuition.

Lambert, on the other hand, in his posthumous work *Theorie der Parallellinien* (1766) argues in a similar way to Saccheri, but does not make the same mistake. In fact, it would seem that he believed it is possible for there to be a geometry without the fifth postulate, as he writes: "From this I should almost conclude that the third hypothesis [of the Acute Angle] would occur in the case of the imaginary sphere." The German mathematician discovered a number of interesting formulae with respect to what are now known as hyperbolic triangles, proving that the sum of the angles is always less than 180°. Lambert's formula states that the following condition holds for such a triangle:

$$(\pi - (\alpha + \beta + \gamma)) = CA_{\alpha\beta\gamma},$$

where:

$\alpha + \beta + \gamma$ is the sum of the angles of the triangle (expressed in radians);

$A_{\alpha\beta\gamma}$ is the total area of the triangle;

C is a constant related to the curvature of the hyperbolic space in which the triangle is situated.

of postulates and hence, a consistent geometry. Both selected the same alternative postulate and proceeded to explore the resulting non-Euclidean geometry, proving theorems in the same way that Euclid had proved his.

The use of this new postulate led to a new system of theorems and corollaries, which is currently known as hyperbolic geometry. The conclusions they reached were unusual. For a given point, there was an infinite number of straight lines parallel to another; the sum of the angles of a triangle was less than 180°; given two parallel straight lines, there was a third that was perpendicular to one of these and parallel to the other…

All this is rather counter-intuitive. We cannot imagine a situation like this without rethinking the concepts of straight, flat, etc. However, from a logical perspective, it is all perfectly valid. This represented a serious mathematical crisis in the 19th century, alongside the other controversies of the period. Yet regardless of this, the work of Lobachevski and Bolyai resolved the dilemma of the parallel postulates referred to in previous sections, and showed that Euclid was correct to include it as a postulate, given that in his geometry it is not a theorem that can be deduced from previous ones.

A note on topology

One of Bolyai and Lobachevski's contemporaries, the German mathematician August Möbius (1790–1868) is best known for the strip that bears his name. Creating the strip is as simple as joining a rectangular band of paper after having twisted one end 180° with respect to the other. 'Walking' over the surface that is obtained, it is possible to cover it in its entirety and return to the starting point without having passed over onto the 'other face', which in fact does not exist. And if we use scissors to cut the strip in half lengthwise in the centre, we obtain not two strips, but one, with a length double the originals.

This amazing result hides a certain trick. According to Martin Gardner, Möbius strips are not, strictly speaking, two-dimensional objects, given that they have a certain width (there is no
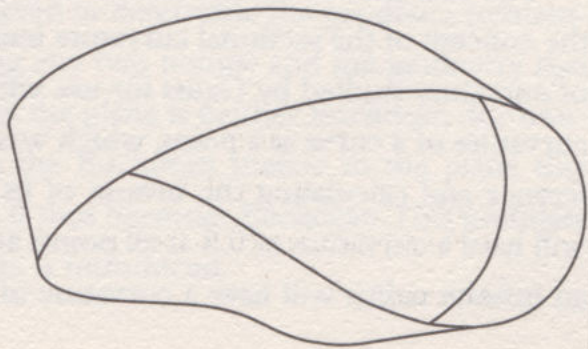
Representation of a Möbius strip, a surface that is connected by 'both sides'.

ribbon or paper whose width is zero). If we now consider the Möbius strip as a three-dimensional object, we can see that its cross section is in fact a rectangle. The strip should be considered as a 'twisted prism'. If this had a rectangular cross-section, before sticking the two bases, we could turn it by just a quarter or half turn (like a traditional strip), or as many times as we like. What would happen if it had a pentagonal section instead of a rectangular one? What object would we obtain? All these objects, and in general the properties of geometric bodies that remain unchanged under various transformations, occupy a branch of mathematics referred to as topology, to which we shall make reference in subsequent chapters.

Flying over Greenland: the model of the Universe

A generation after the discovery of hyperbolic geometry, Bernhard Riemann (1826 –1866), a German mathematician and student of Gauss, gave a lecture at the University of Göttingen in 1851 to complete his training and accede to the post of *privatdozent* (the equivalent of assistant professor). This paper is the most famous one in history to be delivered in order to obtain a university professorship. In it, Riemann sets out a new overall vision of geometry, placing the greatest emphasis on the study of manifolds with any number of dimensions in any type of space. Using a highly intuitive language and without providing proofs, he began by introducing the concept of a differentiable manifold (a generalisation of the concept of a differentiable surface). In fact, the term 'manifold' refers to the different coordinates that can be varied in order to obtain the points of an object, while the adjective 'differentiable' refers to what is smooth, without wrinkles or fractures. According to Riemann's approach, traditional surfaces were two-dimensional manifolds, whereas curves were one-dimensional manifolds and points were zero dimensional. There were also three-dimensional manifolds, however these were not so easy to visualise.

The first part of the lecture culminated in Riemann using geodesics to define the concept of the sectional curvature tensor, which is the extension of the concept of curvature studied by Gauss for use with manifolds. The concept generalised the curvature of a curve at a point, which was measured by tracing the tangent circumference and calculating the inverse of its radius. Thus a circumference of radius 2 will have a curvature of 0.5 at all points and a straight line the tangent of which has an infinite radius will have a curvature of 0 at all points.

This definition is clearly not so easy to generalise for a surface, since in each direction of the plane is it possible to trace a circumference tangential to the surface such that there is an infinite number of tangent circumferences at each point. Which one should be selected? Riemann solved this problem by introducing what is known as the curvature tensor, although he did so not only for surfaces but for any manifold of any number of dimensions.
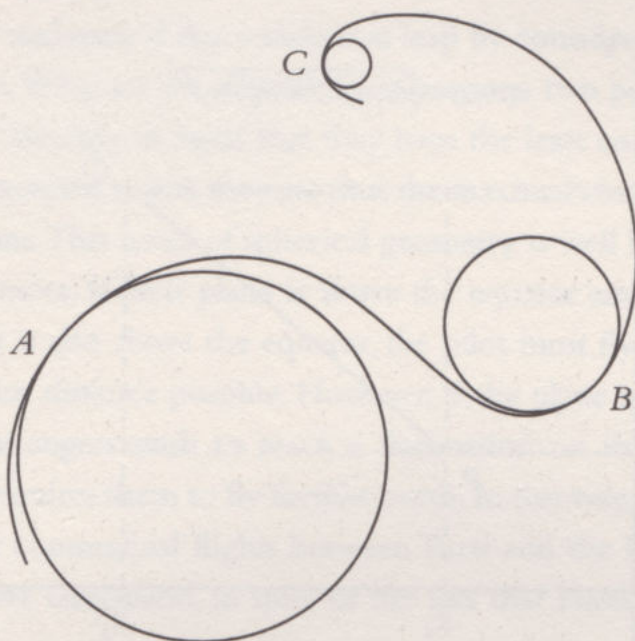
Curvature A < Curvature B < Curvature C

This illustration shows that as curvature increases, the radius of the tangent circumference decreases. (Source: Maria Isabel Binimelis.)

In the second part of the lecture, he considered the model that best explains physical space, the space in which we move. How many dimensions does it have? What is its geometry?

Riemann's new geometric approach considered that any space (be it planar, three-dimensional or any other) can be studied as a differentiable manifold and that introducing a distance or metric to this served to determine the geodesics (remember that these are the shortest lines joining any two points) and the geometry that governed this object. Thus in its own right, the plane is neither Euclidean nor non-Euclidean, and it is only by introducing the Euclidean metric to the plane that Euclid's fifth postulate can be verified and it thus becomes Euclidean. This postulate will no longer be valid if a different metric is introduced.

In the Euclidean metric, for example, in order to calculate the distance between two points based on their coordinates, we begin by drawing the triangle made up of the segment joining the points and the projections of the segments onto those parallel to a given pair of perpendicular coordinate axes that pass through these points. In this way, it then becomes possible to calculate the length of the diagonal using Pythagoras' theorem, as shown in the following diagram:

$$d(P,Q) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$
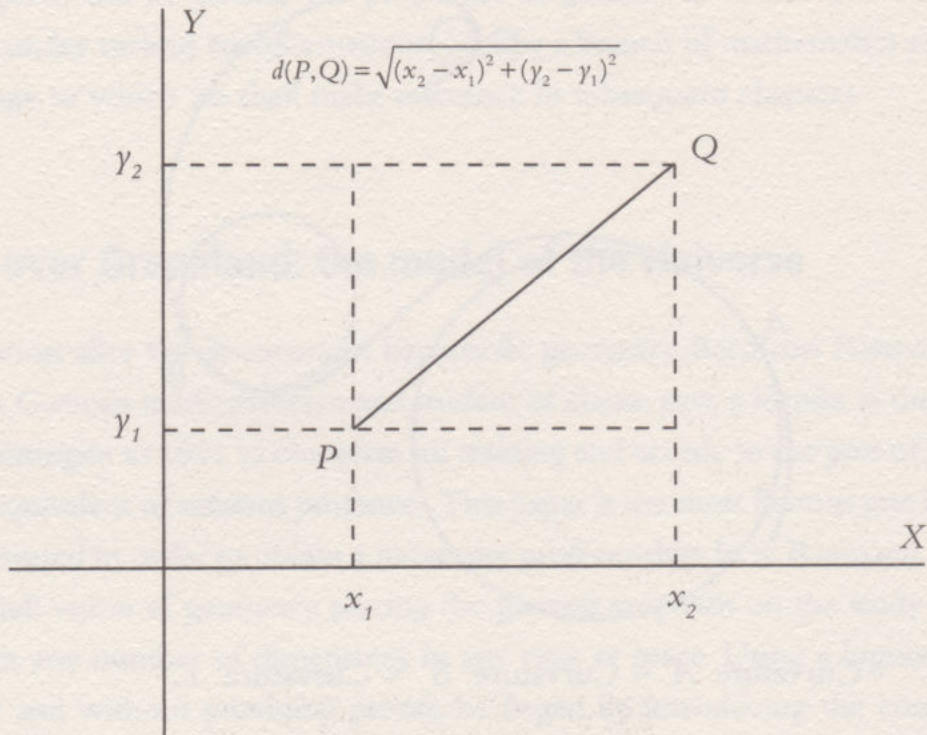
The Euclidean distance between two points P and Q is the hypotenuse of the right-angle triangle obtained by tracing two straight lines parallel to the coordinate axes X, Y, one starting at P and the other at Q, and is calculated using Pythagoras' theorem.

TAXICAB METRIC

Another example of a metric equivalent to Euclidean metric is known as 'taxicab metric', or Manhattan distance, and is given by $d((x_1,y_1),(x_2,y_2)) = |x_2 - x_1| + |y_2 - y_1|$ and measures the distance travelled by a car from one point to another in a grid-iron city (hence its name). Again it can be seen that in its own right the plane is neither Euclidean nor non-Euclidean but instead depends on the metric that is applied.
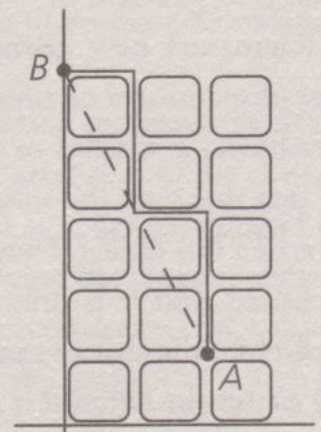
Riemann re-examined the basic assumptions of Euclidean geometry. He analysed the second postulate, "any straight line segment can be extended so as to continue in any direction", and observed that this should be distinguished from "any straight line is infinite". He concluded that this new form of interpreting the second postulate led to him abandoning the fifth. Riemann replaced this by "any pair of straight lines intersect". These considerations lead him to what is referred to today as elliptic geometry.

We can better understand this conceptual leap by considering the geometry of the Earth's surface. What are the shortest lines between two points, or rather, what are the geodesics? Bearing in mind that they have the least curvature, and the least curvature has the greatest radius, they are thus the maximal circles, such as the equator or the meridians. This result of spherical geometry is well known among pilots who fly long distances. If their plane is above the equator and they wish to fly to another point that is also above the equator, the pilot must fly along it in order to cover the minimum distance possible. However, if the plane is at a latitude of 30° north, and the passengers wish to reach a destination on the same latitude, the shortest distance requires them to fly further north. In this way, it is understood that the flight path for commercial flights between Paris and the Hawaiian islands, for example, passes over Greenland, in spite of the fact that Hawaii is to the south of Paris.
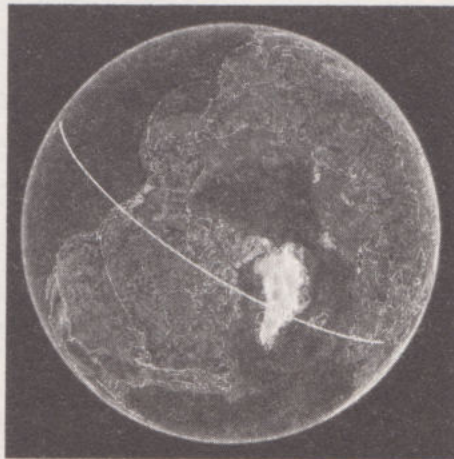

The shortest path between two points is always the geodesic which, in the case of Paris and the Hawaiian islands, passes over Greenland and Canada.

To find the shortest line joining two points on the Earth, we must find the plane that passes through them and the centre of the planet and trace the line that intersects its surface, as shown in the following diagram:
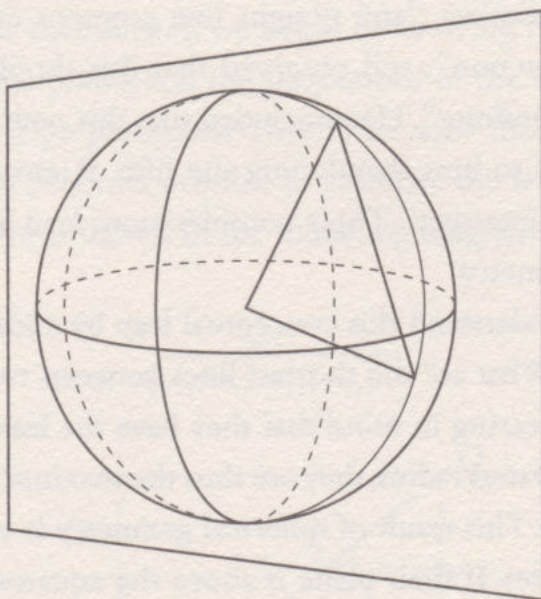
Returning to the idea of parallelism, we can see that this concept does not exist in a spherical geometry in which any pair of 'straight lines' (maximal circles) intersect. On the surface of the globe, we can see that triangles can have two or even three right angles. It suffices to place two vertices on the equator and another on one of the poles. The property that the sum of the angles for all planar triangles in Euclidean geometry is 180° radically distinguishes it from hyperbolic and elliptical geometry. For the latter, the sum of the angles of a right-angle triangle is always greater than 180° and there is no common value for all of them. We can find triangles for which the sum of the angles is 180° and others for which it is 250°. However, it is possible to check that given two triangles with the same area, the sum of their angles is the same.
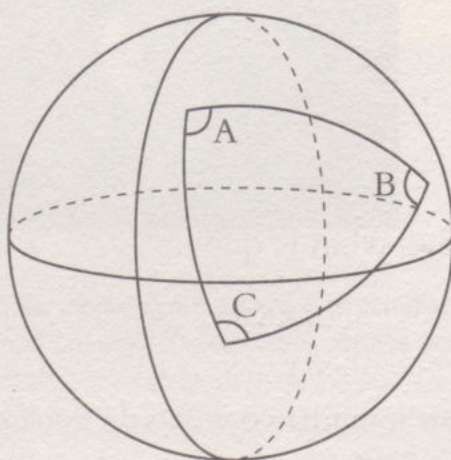
Triangle drawn on the surface of a sphere. The sum of the angles is greater than 180°.

SPATIAL GEOMETRY

Which of the three geometries is the real one? Having reached this point, the question is which of these three geometries best describes the physical world. With time, we have seen that Euclidean geometry is a completely acceptable approximation when working with measurements on an Earthly scale. When the distances are larger, this is no longer so clear. If we attempt to take measurements and travel on the surface of the sphere, we become aware that we live in a world with elliptic geometry. When we travel at speeds that approach the speed of light, the geometry that applies to space-time is Minkovsky geometry, a type of non-Euclidean geometry. But what happens in the space of the Universe that lies beyond the surface of the Earth, without taking time into account? Do we really live in a Universe with Euclidean space?

Gauss spent a period of time working on a geodesic study for the government of Hanover and during the course of this work he had to measure the angles of the triangle formed by the peaks of three mountains located 50 miles apart. The deviation from 180° was less than the estimated error for the measurement, meaning that the calculation was consistent with the three hypotheses. For his part, Lobachevski observed that a triangle located on the Earth was too small to allow the differences to be seen. As a result of this, he proposed to study an astronomical triangle... However this failed to lead to a conclusion because the differences in Sun-Earth distances are less than one thousandth of a second. In turn he began to consider larger triangles and looked to the parallax of the stars. But neither Lobachevski nor anybody else has been able to find a triangle in which the sum of its angles is not 180°, in spite of the fact that in hyperbolic geometry, the larger the area, the larger the difference should be. At present, according to the theory of relativity, Riemann's elliptic geometry best fits with the metric of the Universe. In the words of B. Lewis: "In Einstein's general theory of relativity, spatial geometry is a Riemannian geometry. Light travels along geodesics, the curvature of space is a function of the nature of the matter of which it is composed."

The Erlangen Programme: what is geometry?

Upon joining the Faculty of Philosophy at the University of Erlangen, Felix Klein (1849–1925) wrote an unpublished paper in 1872 which, together with Riemann's lecture and Euclid's *Elements*, can be considered as one of the essential texts to be studied in geometry. In it, Klein tackled the problem of providing a formal definition of geometry, beyond the intuitive idea that we may have of it. The

systematisation of the multiple geometries that had been suggested up to that point was presented in what is known as the Erlangen Programme. This classified each geometry according to the properties that remained constant with respect to a certain group[12] of transformations. The concept of a group was not invented by Klein although it was he who discovered this fundamental relationship between geometry and the group of transformations. Thus, Euclidean geometry is the study of the properties or quantities that remain constant – or invariant – for the group of congruent transformations. Congruent transformations or isometries (which in Greek means 'equality of measure') are translations, symmetries, rotations and combinations of these three. Their invariants are, for example, the distance between two points, the areas of surfaces, the angles between straight lines, etc.

Likewise, affined geometry is the study of invariants (in this case colinearities) with respect to the group of affined transformations (isometries together with scalings and shears); projective geometry is the study of invariants with respect to the set of projective transformations; and topology is the study of invariants with respect to the group of continuous and inverse continuous functions.

Among many other things, Klein proved that Euclidean, affine and non-Euclidean geometries could be considered special cases of projective geometry. Broadly speaking, he achieved this by considering transformations on projective space that leave a certain conic (the absolute conic) invariant. The type of geometry we are left with depends on the type of conic.

Technicalities aside, this last point implies something that is extremely important. Euclidean geometry is consistent (that is to say, it does not have contradictions) if and only if this is also the case for non-Euclidean geometries. This ended the controversy of whether non-Euclidean geometries were coherent, although the matter lingered on for a number of years in light of scepticism voiced by those who considered Klein's argument to be erroneous.

---

12 For a set with an operation to qualify as a group, it must meet certain conditions: (i) the operation must be associative, meaning that given any three elements, a,b,c, from the set, the result of applying an operation to the first two (a and b) and applying the operation to the result and the third (c) must be the same as applying the operation to the first (a) and the result of applying the operation to the second and the third (b and c); (ii) there must be an identity element, that is to say an element e belonging to the set such that taking any other element of the set a and applying the operation in both directions, the result is still a; *(iii)* finally, each element must have an inverse, or rather, for any element a of the set, there is an element a^{-1} such that applying the operation in both directions gives the identity element.

The Programme paved the way for the study of abstract geometric spaces. Analysis was no longer limited to figures on the plane or the three-dimensional space in which we move. Now it was possible to think in many dimensions using variables that were not necessarily special. Thus, for example, we can talk of the space of the variables of the thermodynamics of a gas, which can include more than three, such as pressure, volume, temperature and the various concentrations of substances that make up the gas. Such spaces can now be used to study geometric properties from a more abstract perspective.

A grain of pollen and the geometry of nature

The shapes studied by Euclid had many limitations when it came to explaining the Universe. The geometry of nature is far removed from the perfect figures imagined by the Greek master.

At the beginning of the 19th century, the Scottish botanist Robert Brown discovered a droplet that had become trapped in a fragment of igneous rock as it had solidified. Focusing on this droplet under the microscope, he observed minuscule traces of particles that oscillated in a completely irregular motion without coming to a halt. The movement was familiar to him as he had already observed it when studying grains of pollen suspended in water. Brown's explanation of the phenomenon was that the vitality of plant molecules is retained for a while after their death. However, this explanation was subsequently discredited and people began to believe that the agitation, known as Brownian motion, was caused by a physical phenomenon rather than a biological one. For example, the smaller the particle or the greater the temperature, the faster its movement.

It was not until 1905 that Albert Einstein established the first implications of the kinetic theory of gases devised by James Clerk Maxwell and Ludwig Boltzmann, applying them to the study of Brownian motion. The phenomenon is now explained in the following manner. A grain of pollen suspended in a liquid is bombarded by the molecules of the liquid and their impact on the grain causes its path to alter constantly. On the one hand, the fluctuations occur arbitrarily and on the other, since the microscope only shows fluctuations of a certain size, the movement that is observed can only hint at the complexity of the actual path.

The trajectory of a particle in Brownian motion was one of the first natural phenomena in which self-similarity was recognised on different scales. The

following drawing shows Brownian motion as represented in 1912 by the French physicist Jean Baptiste Perrin, who recorded the position of the particle every 30 seconds.
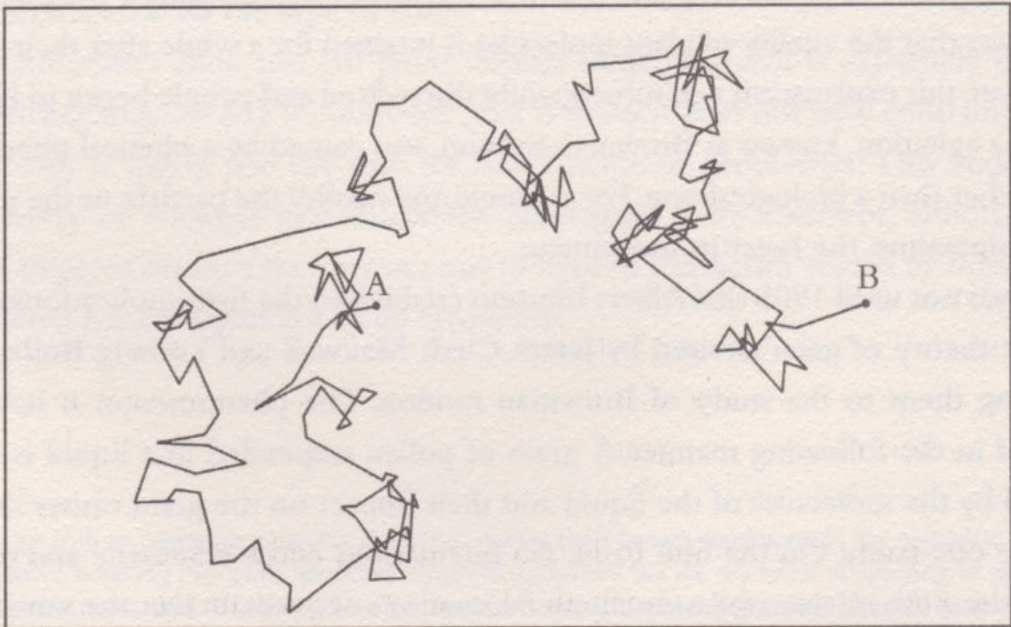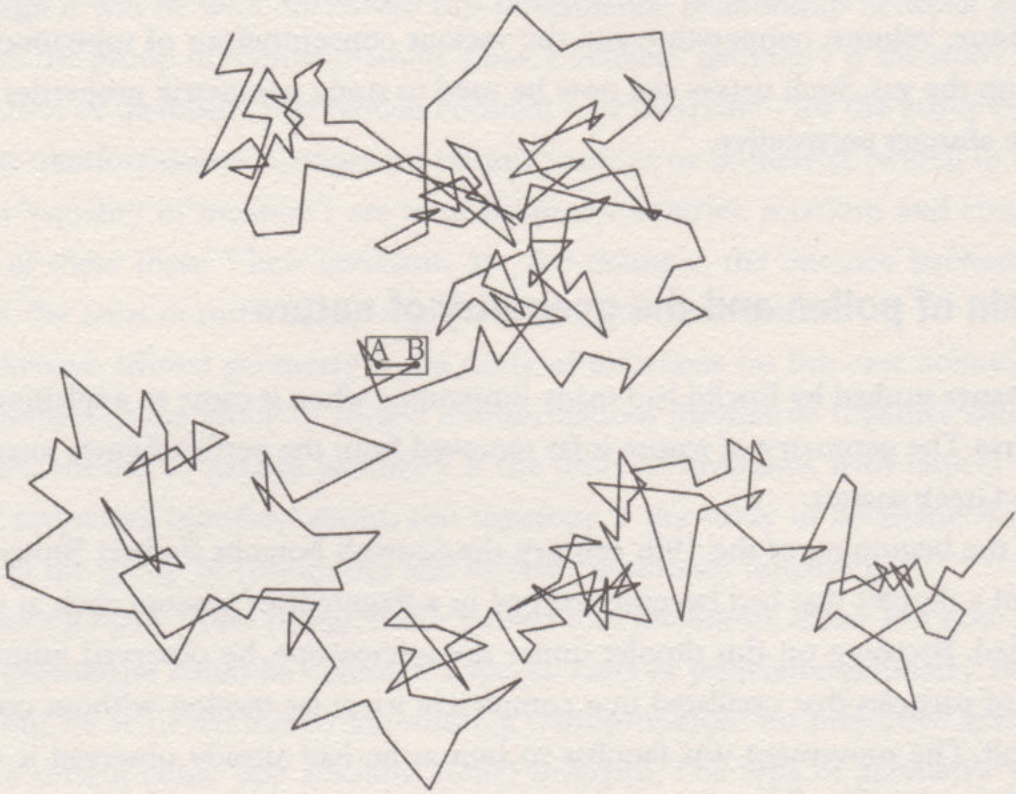
Graph of the Brownian motion of a particle showing both the complexity of its trajectory and its similarity at different scales. The bottom image shows a magnified representation of the path between the points A and B in the top image.

The lines of the path of a Brownian particle do not, strictly speaking, have any physical reality. The lines in the drawing are traced between the positions of the particle of pollen recorded every 30 seconds, marked with a point and then joined up using straight lines. As such, the only real elements are the points that indicate the positions of the Brownian particle at the end of each interval. Thus if, for example, we focus on two points in the sequence, A and B, and now, instead of marking the positions at 30-second intervals, we do so every 0.3 seconds, joining the points with straight lines, as before, they are replaced by a series of smaller but equally complex jagged lines. We can now take even smaller intervals, such as 0.003 of a second, and follow the same procedure. The same thing happens as before. It becomes clear that the path followed by a Brownian particle is such that it maintains a pattern, the complexity of which remains similar as we change the intervals at which is it observed.

It is interesting to note that Perrin was already aware of this type of behaviour in 1906. In particular, he had noticed that by taking a point from the path followed by a Brownian particle, it is impossible to strictly trace a line tangential to it and noted:

"In geometric terms, curves without a tangent constitute the norm, and regular curves, such as the circle, are interesting but special.

[…] Those heard talking of curves without a tangent tend to think that nature does not give such complications, and nor does it suggest them.

We are within the realm of experimental reality when we observe the Brownian motion with which a [Brownian] particle moves when suspended in a liquid under the microscope. We then discover that the direction of the straight line joining the positions occupied by the particle at two points extremely close in time, varies irregularly in an absolute way as the interval between the two moments is decreased. As a consequence, an unprejudiced observer would conclude that they are dealing with a curve for which it is not possible to draw a tangent."

No one paid any attention to Perrin's comments and the subject lay dormant until the end of the 1960s when the French-American mathematician Benoît Mandelbrot took it up. If the observation made by Perrin at the start of the century had continued to be investigated, it is almost certain that the foundations for a new geometry would have been developed six decades earlier.

Benoît Mandelbrot was born in Poland in 1924, to a Lithuanian Jewish family. He emigrated to France in 1936, were his uncle, Szolem, a founding member of the Bourbaki[13] group, had settled. This school of thought explicitly rejected the use of geometric or graphical figures in order to illustrate concepts or proofs. The eyes, they said, could trick the mind.

In 1945 his uncle recommended that he read a 300-page work by the French mathematician Gaston Julia, entitled *Memoir on Iteration of Rational Functions*, and as was to be expected from a member of his mathematical school, he added: "Forget the geometry." Mandelbrot ignored this second recommendation, and did not take much interest in reading the book until 1970 when, with the help of the computers provided by IBM for the Thomas J. Watson Research Centre, he contributed to the creation of illustrations of an experiment that surprised the scientific community with the detail of an image that was later to be called the Mandelbrot Set.

Together with other mathematicians who preceded him, Mandelbrot led the rise of a new way of thinking in mathematics and the natural sciences which, on account of his efforts and creativity, became an event of the first order: fractal geometry.

"Fractal geometry will make you see everything differently. There is danger in reading further. You risk the loss of your childhood vision of clouds, forests, galaxies, leaves, feathers, flowers, rocks, mountains, torrents of water, carpets, bricks and much else besides. Never again will your interpretation of these things be quite the same."

Thus begins the book *Fractals Everywhere* by the English mathematician Michael Barnsley, professor at the Australian National University and prestigious researcher in this field. According to Barnsley, strictly speaking, fractal geometry is a new language. Continuing with the linguistic analogy given in the earlier section on urban development, and making use of a metaphor devised by the researchers Hartmut Jürgens, Heinz-Otto Peitgen and Dietmar Saupe, let us begin to tackle some of the fundamental properties of this geometry.

---

13 Nicolas Bourbaki is the collective name of a group of French mathematicians from the 1930s, which proposed the revision of the fundamentals of mathematics under extremely rigorous criteria. Its members included illustrious figures such as Henri Cartan, Jean Dieudonné, André Weil, Jean-Pierre Serre and Alexander Grothendieck.

While Western languages are written using a finite alphabet (for example, the Latin alphabet), Eastern ones, such as Chinese, make use of a large number of symbols. In the former, it is necessary to combine letters to make up words, which have meanings, whereas in the latter the symbols already have a meaning in their own right. Just like Western languages, traditional geometry (e.g. Euclidean or Riemannian) involves finite elements, such as the straight line or circumference, which are used to construct other, more complex shapes, whose meaning depends on the context in which they are included.

In contrast, fractal geometry corresponds to the family of Eastern languages in the sense that it works using elements with a meaning in their own right (in this case infinite), in contrast to those of traditional geometry. What then are these elements? The simplest way is to identify them with rules for their calculation, or algorithms, which can be considered as units of meaning in the fractal language. An algorithm is a set of rules and instructions for a procedure that often requires the help of a computer to produce concrete shapes and structures.

From this perspective, classical geometry is a first approximation to the structure of physical objects. In fact, differential geometry offers an excellent approximation of such objects. As an example, an observer on the Earth can state that the sphere is a suitable model of the Moon. However, for an astronaut who finds himself on the moon observing the various craters, this would not be a valid model. Thus, using traditional techniques to model the complex and irregular structures that surround us is extremely complicated. In a certain sense, fractal geometry fills this gap and can be used to faithfully design everything from the silhouette of a leaf to the growth of the tree on which it grows.

It is difficult to provide a general definition of the term *fractal* because many of these cannot be applied to all the families of fractals that exist. Perhaps the best way of describing them is to note what is common to the mathematical processes by which they are generated. At the end of the day, the most interesting fractals and the root of their deepest mathematical properties can be found in the characteristic structure of the processes in which they are rooted.

Thus put, a fractal is a final product that arises from the infinite iteration (that is to say repetition) of a well-defined geometrical process. This elementary geometrical process, generally of an extremely simple nature, determines the final structure which, due to infinite repetition, gives rise to structures of extraordinarily complex appearance.

FUNDAMENTAL DIFFERENCES BETWEEN EUCLIDEAN GEOMETRY AND FRACTAL GEOMETRY	
Euclidean	Fractal
Traditional (over 2,000 years)	Modern (approx. 50 years)
Whole dimension	Fractal dimension
Deals with man-made objects	Suitable for naturally occurring forms
Described using formulae	Described using recursive algorithms (iteration)

In fractal geometry, the process responsible for an intricate and complex phenomenon can be surprisingly simple. And the converse is also true: the simplicity of a process should not lead us to underestimate its possible consequences, which can often be highly complex.

The essence of Mandelbrot's message is that many natural structures (such as mountains, clouds, coasts and capillaries), which appear to be of extraordinary complexity, in reality present the same geometric regularity – their invariance at different scales.

In 1975 Mandelbrot published an essay entitled *Les Objects Fractales: Forme, Hasard et Dimension (Fractals: Form, Chance, & Dimension)* which explains that the neologism *fractal* comes from the Latin *fractus* ("broken" or "fractured"). Later on, the American popular science writer, James Gleick recounts in his book *Chaos, Making a New Science* that one afternoon, while preparing to publish his work, Mandelbrot thought he should give a name to his shapes, dimensions and geometry. His son had returned home from school, and Mandelbrot began to read his Latin dictionary. He came across the adjective *fractus*, derived from the verb *frangere*, 'break', and the resonance of the first related English words – *fracture, fraction* – struck him as apt. In this way, he created the word *fractal* making it both a noun and an adjective.

In 1982 he published a new book that included spectacular images created using computer technology: *The Fractal Geometry of Nature*. On page 15 of the first edition of the work, Mandelbrot proposed a definition that he himself acknowledged did not cover sets which, for other reasons, should be included under the category of fractals: "A fractal is a set for which the Hausdorff–Besicovitch dimension strict-

ly exceeds the topological dimension."

Other definitions have been proposed and in fact we are dealing with a geometric concept for which there is still no precise definition, nor a single and commonly accepted theory.

Mandelbrot did not invent fractals but was in the right place at the right time to make their discovery and realise their secrets. They have been humankind's invisible companions since the beginning of time, just like chaos, which will come to be the invisible hand that rocks the cradle. Mandelbrot died in Cambridge on 14 October 2010.

Chapter 2

The Unknown Dimension: Mapping the Universe

So, naturalists observe, a flea
Has smaller fleas that on him prey;
And these have smaller still to bite 'em,
And so proceed ad infinitum.
Jonathan Swift, *Poetry, a Rhapsody* (1733)

In Holland in 1904, boxes of cocoa powder appeared with an image that contained a most peculiar repeating element. It was not the first time that an artist had made use of this effect. Many centuries before, in 1320, Giotto had used it in an altar triptych painting commissioned by Cardinal Stefanschi. However a long time passed before it attracted anyone's attention. It was in 1970 when a Dutch journalist wrote an article on the topic and coined the term the 'Droste effect', making reference to a brand of cocoa powder.

A Universe in a drop of water

The box in question contains an image of a nurse holding a tray with two objects. These two objects attract our attention precisely because the same nurse in the same pose is contained within them and this pattern is repeated to the limits of our perception. If we were to place ourselves within one of these labels, we would see everything as if we were outside. In fact, we could only know which label we were in if the scale of our body did not change.

We find ourselves with a case of partial self-similarity. The term self-similarity is used because the small images are similar to the whole, and partial because the large image is not completely made up of small nurses. In order to achieve full self-similarity there would need to be a crowd of nurses that covered everything when we zoomed into any part of the image.

The recursive image on this pack of cocoa powder gave rise to the term the 'Droste effect'.

Let us now focus on a well-known cartoon: The big fish eating the small one. At first sight, it would seem to exhibit the same sort of visual repeat as the boxes of cocoa powder. There are an infinite number of fishes, each behind the smaller one which it is going to eat. However in this case, if we expand any part of the figure, we can see that the scale of any fish is full of small fishes eating each other. Here we have a case of full self-similarity achieved using eleven similarity functions. Each function converts the larger shape into a smaller one which is rotated and/or moved, fitting within the larger frame. In this way, at the first step, the large fish is covered by eleven smaller fish. With the image we have generated, we can continue to infinity in such a way that in the end the collection overlaps like a collage.

This type of function was discovered by Barnsley, who proved the collage theorem. Further on, we will go into details of how such functions are created and discover how the procedure 'works' thanks to this theorem.
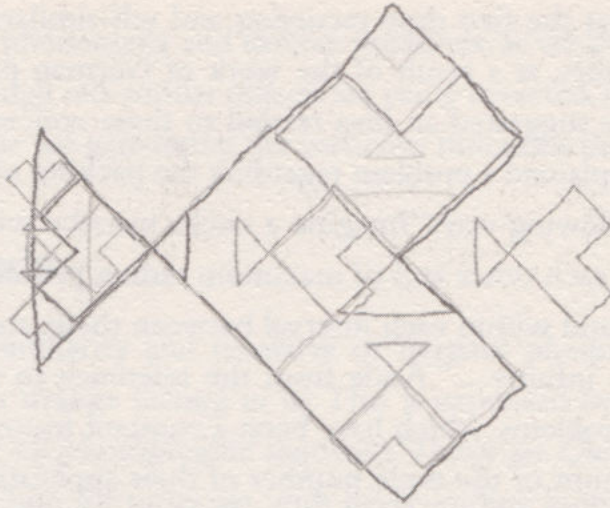
First iteration of the pattern of the large fish eating the smaller one.
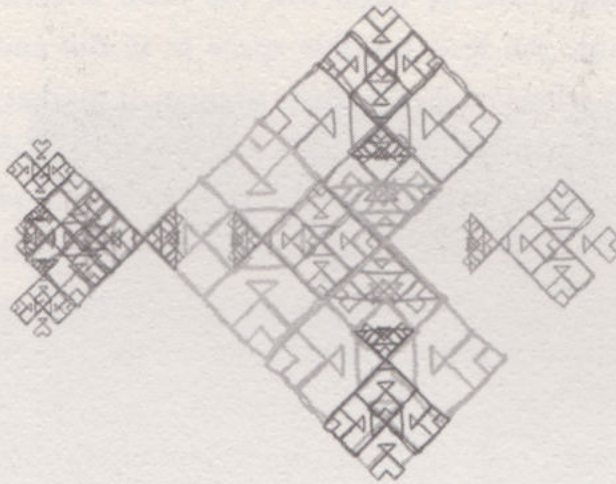
Third iteration of the pattern.
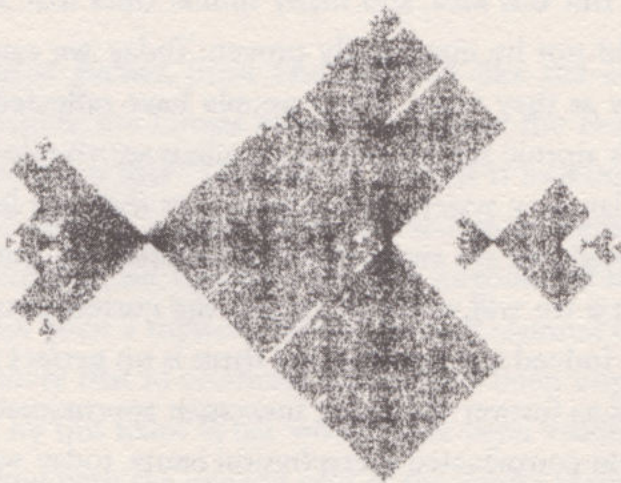
Final iteration of the design.
(Source of illustrations: Maria Isabel Binimelis and Laura Elisabeth Violant.)

It would seem that the first time recursion and self-similarity explicitly arose was in the 17th century, as a result of the work of German polymath Gottfried Wilhelm Leibniz. He suggested an idea related to these concepts on at least two occasions. Leibniz explained a problem regarding the packing of geometric figures to a friend in the following way: "Imagine a circle; inscribe within it three other circles congruent to each other and of maximum radius; proceed similarly within each of these circles and within each interval between them, and imagine that the process continues to infinity…" Aside from the reference to recursion, Leibniz' interest in packing problems, which have been a constant feature in the history of mathematics on account of the wide number of their applications, is also present in this excerpt.

Perhaps the most famous packing problem has been the conjecture proposed by Kepler. In this conjecture, he states that the most effective way of packing cannon balls to occupy the least possible space is, in the end, to use the same method used by fruit sellers to display their oranges at market stalls. But such an apparently simple conjecture was not fully proven until 2005 with the help of computers. The majority of packing problems are rooted in physics and biology and are motivated by a large variety of applications in crystallography, the structure of amorphous materials and colloid aggregates. Even the optimal transmission of digital signals can be formulated as a version of the sphere packing problem, known as the 'kissing number' problem.

On a second occasion, Leibniz stated that the entire Universe was contained in a drop of water, which in turn contained smaller drops of water, each of which contained the entire Universe which contained other drops of water even smaller, and each of these… But this idea, and many similar ones that arose, were in time rejected as they could not be empirically proven. Today we can confirm that the ideas are not as crazy as they seem. Many people have reflected on the similarity between Niels Bohr's atomic model, with its nucleus surrounded by orbiting electrons, and Kepler's planetary system. It is a similarity that also links the micro and macro cosmos. Although we are aware that these models are not completely accurate, it is also clear that we will never be certain the current model is the one that best describes reality. Indeed it is possible that there is no perfect model and that we must resign ourselves to forever obtaining successive approximations.

Leaving aside these complicated metaphysical issues, today we know there are a large number of objects that contain themselves. Some of these have been theoretically described in mathematics; others appear in nature. Self-similarity appears

in a wide range of phenomena and diverse situations, as we shall see in subsequent chapters. And although self-similar objects are easily detected, in practice there is no automatic technique for generating the recursive functions that can create them.

Infinity in a circle

The fact that self-similarity and recursive or iterative phenomena are so closely linked can arouse a strange feeling in us. This is expressed by the American artist Duane Michaels in his photographic sequence *Things are Queer*. The human mind feels a special attraction for processes with never-ending loops; they lead us to experience the concept of infinity, as intractable as it is seductive. Perhaps unconsciously, perhaps not, artists often offer as much as scientists in this quest.
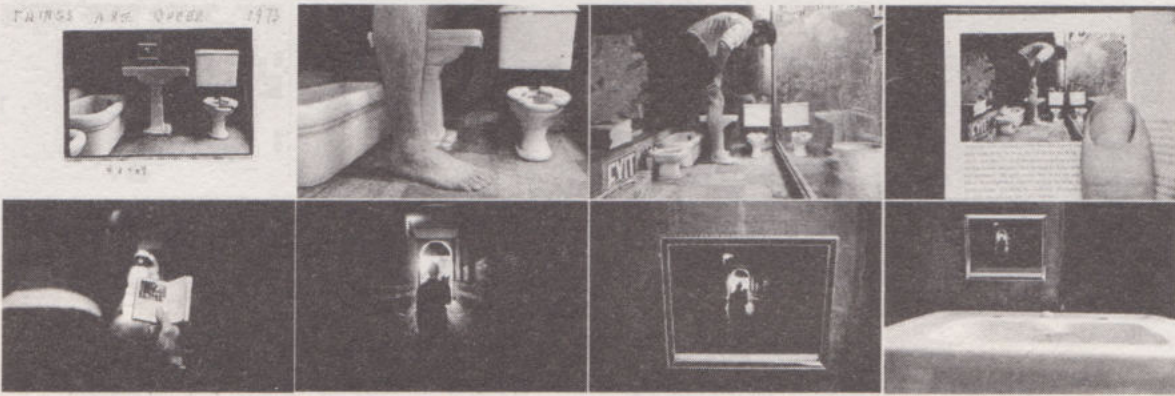
Sequence from Things are Queer, *by Duane Michaels.*

The artist Maurits Escher, from Holland, just like the cocoa powder boxes, shows us, albeit invisibly, the Droste effect in his piece the *Print Gallery*. Referring to this work, Escher states that "The boy on the left is looking at a print in which he himself appears." But we cannot verify this statement because just at the point where the boy is looking, his signature appears surrounded by a white halo. This blind spot has always been a mystery. Escher himself claimed that "Everything becomes so detailed there that to continue would have been impossible." What is the mystery concealed by this halo? What would have been visible if Escher had continued drawing in line with the general grid of the work? Here the term grid refers to the network of lines that correspond to the horizontals and verticals of the drawing without deformation. For example, the lines that make up the 'vertical' sides of

the frames of paintings and windows make up one part of this grid and the 'horizontal' ones the other.

What is really hidden behind the first arch only came to light in 2000, when fellow Dutchman Hendrik Lenstra analysed this structure. First of all he reconstructed the *Print Gallery* without the distortions and upon doing so, a white area in the shape of an infinite spiral was revealed in the middle of the drawing. He then completed this area based on the existing structure.
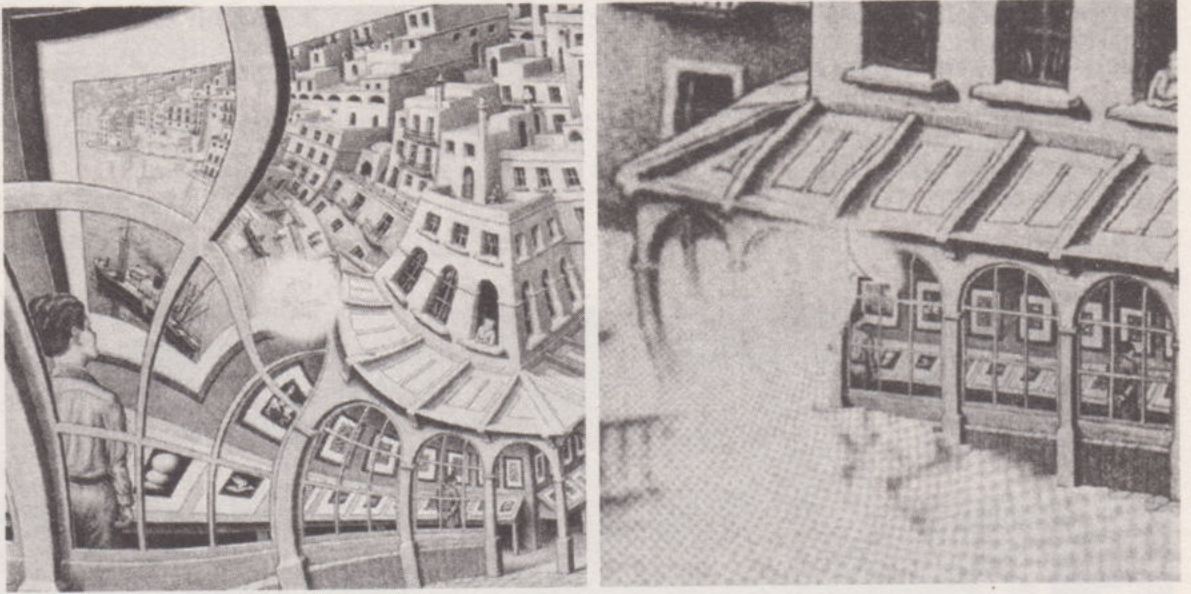
Detail of Print Gallery, *on the left, and the straightened version.*
(Source: 'Notices of the AMS', Volume 50 Number 4.)

Bruno Ernst, a scholar of Escher's work described the succession of enlargements of the straightened version thus: "A boy is viewing Escher's prints in an art gallery and his gaze falls on a print that is extremely similar to that of *Malta* (1935). The boy is paying particular attention to the top right corner where he sees part of the Senglea peninsula. There he discovers a building with a soffit that reminds him of the gallery into which he has just entered. Focusing in a little more detail, he observes that there are works by Escher in the gallery, and he can even see a print of a boy looking at a print of Malta."
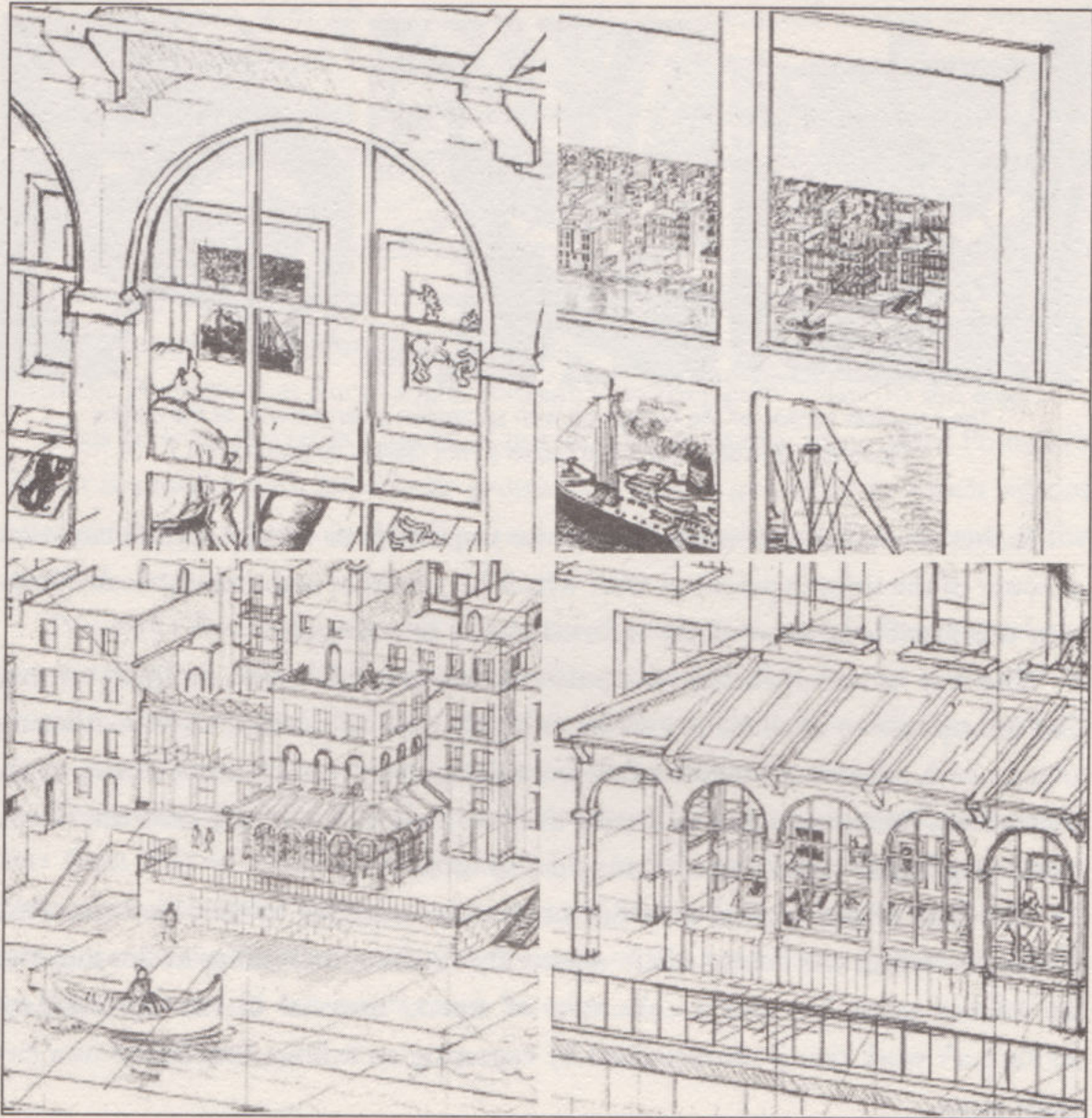
Successive enlargements of the straightened drawing.
(Source: The Droste Effect and Escher's Print Gallery *by Bruno Ernst.)*

Based on the mathematical description of the structure and the straightened drawing, Lenstra created a computer programme that allowed him to expand the drawing towards the inside of the circle, far beyond the border where Escher had stopped… towards infinity. In this way, the blind spot disappeared and the 'corrected' version of the *Print Gallery* appeared, infinitely repeating itself in its own centre. The most curious thing is that the information that was initially completed in the spiral was not necessary to continue the grid to the centre.
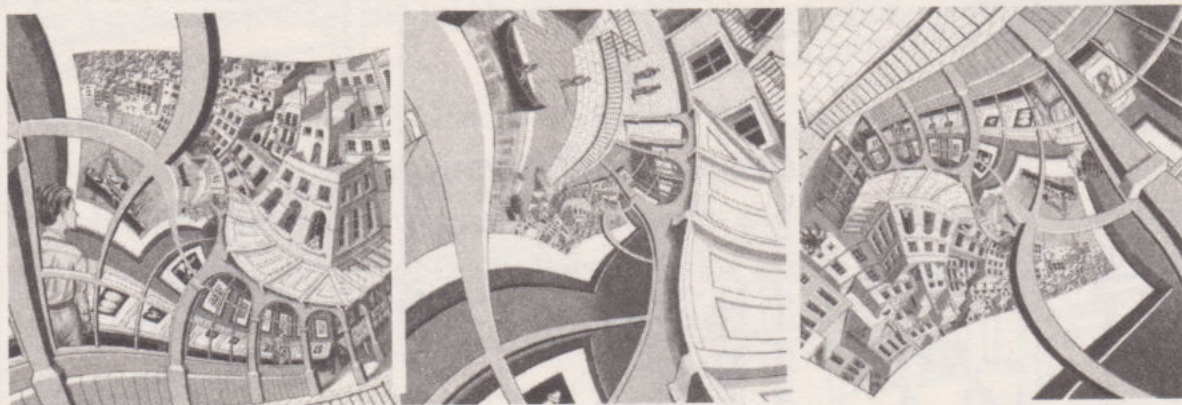
The complete version of the lithograph with successive enlargements of the centre.
(Source: The Droste Effect and Escher's Print Gallery by Bruno Ernst.)

Escher was correct in saying that it was impossible to fill the central circle on account of the infinite precision with which he would be required to work. However it is possible to show all the information in an animation.

The way in which Escher manipulates space, generally in an intuitive manner and in response to aesthetic principles, is highly unusual in the art world. Altering objects with reflections, stretches, deformations, projections and other similar transformations, Escher conceals universes that are more complicated than the norm, exhibiting highly sophisticated mathematical tools.

The title of this section does not only make reference to the work the *Print Gallery*, in which the central circle repeatedly contains everything in the shape of an infinite spiral, but also a collection of aspects essential to the development of Escher's work that is characterised by capturing an infinite tessellation in a circular space.

In fact, it is possible to make out three essential types of print in the artist's work. Firstly, there are those that show a landscape or everyday scene from unusual perspectives, thus giving them an intriguing appearance. Secondly, there are those that recreate impossible architectures or shapes. Finally, the third type covers prints in which the plane, or part of it, is fragmented.

The coverage of a surface based on tessellations created by a series of symmetries and periodicities gives us a sensation of infinity on account of their indefinite extension. Escher, who was obsessed with visually capturing the infinite, was not wholly satisfied with this type of representation, since in practice this capacity for infinite extension is a sham, and the work must end at a border. Escher sought to represent infinity in a finite space, leading him to create various works in which circles and squares were linked together.

He continued with his research in which he represented a series of repeated figures fitted inside each other. This time he started at the outside of a square or round border and worked inwards so that as they approached the centre they became smaller and smaller. However, it would appear that he was not wholly satisfied with this idea.

In 1958 he read an article by the British scientist H.S.M. Coxeter entitled *Crystal Symmetry and Its Generalizations*. In it he found a tile, which in some senses inspired a change of direction in his investigations. It was a triangulation of the circle with the property that the triangles multiplied as they grew closer to the edge.

This drawing describes a model of non–Euclidean metric space called Poincaré's disc. It is a model with hyperbolic geometry, a type of geometry in which various lines parallel to a given straight line pass through a given point, as we saw in the previous chapter.

Poincaré's disc makes up part of a series of important models from hyperbolic geometry, given that there is no three-dimensional surface in real space ($\mathbb{R}^3$ in mathematical notation) that fully describes it[1]. As such it is very different from elliptic geometry in which the sphere is used as the model for study.

The model described by Poincaré consists of a circle with a metric that differs from that of the Euclidean plane. The metric property that characterises Poincaré's disc is the one viewed from outside, everything grows smaller as we get closer to the circular border[2]. As a result of this property, a person who lives in Poincaré's world can never reach its edge.

---

1 In more precise technical terms, we say there is no complete (any geodesic can be infinitely extended) isometric (which conserves distances) embedding of the hyperbolic plane in $\mathbb{R}^3$, although there are local embeddings, such as the pseudo-sphere, and also embeddings where the first derivative is continuous.

2 In this metric, the distance between two points P and Q is defined as $d(P,Q) = \left| \ln \frac{PA \cdot QB}{PB \cdot QA} \right|$, where A and B are points on the border of the disc that intercept the only straight line passing through P and Q. (PB denotes the normal Euclidean distance between P and B, and similarly for PA, QB and QA.) This definition, thanks to the work of the eminent British mathematician Arthur Cayley, verifies the properties of a metric and perfectly models the idea that the distance increases as we move closer to one of the points on the border of the disc, with the distance of any point on the border finally being infinite.

HYPERBOLIC AND UNCONSCIOUS

Would the beings living in Poincaré's world be aware that they lived in that space? Let us imagine that one of those inhabitants measures the length of the palm of their hand and notes that it is 20 cm. After walking for a while in the direction of the edge of the disc, they measure it again. We would see that it has shrunk. They, on the other hand, would see that the length of the palm continues to measure 20 cm since the metric belt has also shrunk. The measurements are relative. Whereas for us, looking in from outside, it appears that the palm has shrunk, it remains the same for the person living on the plane. Similarly, for us, their world is enclosed, however for them it is unlimited as they will never reach the edge. How can they discover that they live on a hyperbolic plane? One possible way would be to add together the angles of a triangle to show that their total is less than 180°. This system will always work provided that the triangle is large enough to compensate for the error in measurement, since as the size of the triangle increases, the sum of its angles will decrease. Another way would be to trace a circumference with radius r and check that it measures more than $2\pi r$. But here the radius of the circumference must also be of a suitable size.

In his *Circle Limit* series, Escher specifically attempted to visualise this metric and the appearance of the straight lines in this model of hyperbolic geometry while at the same time giving his own interpretation of infinity: The universe in a drop of water. Different configurations of tessellations can be used. The one that appears in
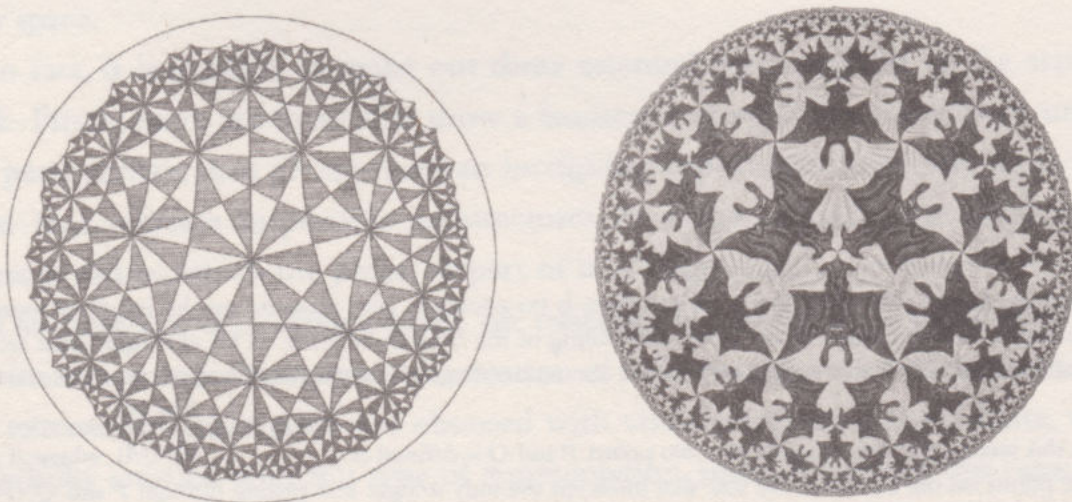
On the left, a triangular function that appears in a work by Poincaré on elliptic functions which, as the author would go on to explain, made use of non-Euclidean geometries. On the right, Escher's Angels and Devils.

the drawing opposite, used by Poincaré in one of his pieces of research, consists of heptagons, each of which is connected to two more heptagons at each of its vertices. The configuration of Escher's work *Angels and Devils*, in which the pentagons with five 'right' angles connect to three more at each vertex, is particularly interesting.

Wars and border lengths

Jaromir Korcak, a Czech geographer and statistician studies the influence of the geographical environment on the population. In 1938 he carried out a number of statistical studies on the number of large islands in certain regions throughout the world, which led him to deduce a law that is essential for the development of the mathematical concept of dimension. Given a certain area, *s*, he calculated the number of islands with a surface greater than *s*. In this way, for each *s* he counted the number of islands with these properties, *N(s)*, and plotted the points on a coordinate axis. He repeated the procedure for different regions, thus obtaining a graph of points associated with each. He observed that the graphs had something in common: *N* depended on an inverse power of *s*, or rather, *N* is a certain constant *k* divided by *s* to a certain power, which we shall refer to as *D*:

$$N(s) = k/s^D.$$

Hence he was able to state a value of *D*, which characterised each region he studied. His observations have now been improved, and we know that the *D* associated with Africa (where a large island dominates other smaller ones) is 0.5; the number associated with Indonesia and North America (where large islands are less predominant) is 0.75; and the number associated with the planet in general is an average of 0.65.

During the last century, a range of studies have developed laws similar to Korcak's although related, for example, to the range of vocabulary of people or the river levels of the Nile. All these have also been described using a certain exponent similar to *D*. However, the most notable law of all is doubtless the one developed by Lewis Fry Richardson (1881–1953), an English scientist and pacifist who pioneered the application of modern mathematical techniques to the prediction of the weather and similar ones to the study of the causes of wars and how to prevent them. While immersed in the latter research, he decided to explore the relationship between the probability of two countries going to war and the length of the border they shared.

HOW LONG IS A SPIRAL?

Spirals are a type of object that challenge the traditional measurement of length. A timeless object of fascination for mathematicians, Archimedes wrote a treatise on them and gave his name to a type of spiral. Archimedes' spiral is like the section of a carpet that has been rolled up, meaning that the length between each turn is constant. The formula that describes it is as follows: $r = q \cdot \phi$, where r is the length of the point of the spiral as a function of the angle ϕ between this point and the x axis, measured in radians and in an anticlockwise direction, and where q is a constant that gives the distance between successive turns when multiplied by 2π. Logarithmic spirals are another type of spiral with an appearance as follows.
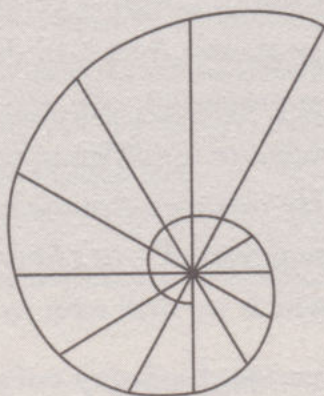
They are obtained when the product of the coefficient q and angle ϕ gives the logarithm of r instead of r. The Swiss mathematician Jacob Bernoulli was so impressed by the similarity of a part of this spiral with the whole shape (its self-similarity), that he chose the following inscription for

Richardson spent many years collecting data on the borders between countries although the results of these studies were not published until 1961, eight years after his death. In his article, he noted that different countries gave different measurements to the borders they shared (to discover this all he needed to do was consult different encyclopaedias). Spain, for example, claims that its border with Portugal measures 987 km, whereas for Portugal the figure is 1,214 km. Similarly, Holland claims that its border with Belgium is 380 km, whereas the latter states it to be 449 km.

Richardson attempted to use graphs to investigate why these huge differences, often as much as 20%, arose. His explanation, as obvious as it was extraordinary, was that the unit of measurement used by one country could be much smaller than that used in another. What were Richardson's experiments? Assume a surveyor opens their

his gravestone: *Eadem Mutata Resurgo*, which can be translated as "I will rise again the same yet changed." In a stricter sense however, the property admired by Bernoulli is that the growth of the spiral is the same as from its centre, rotating it by a certain angle.
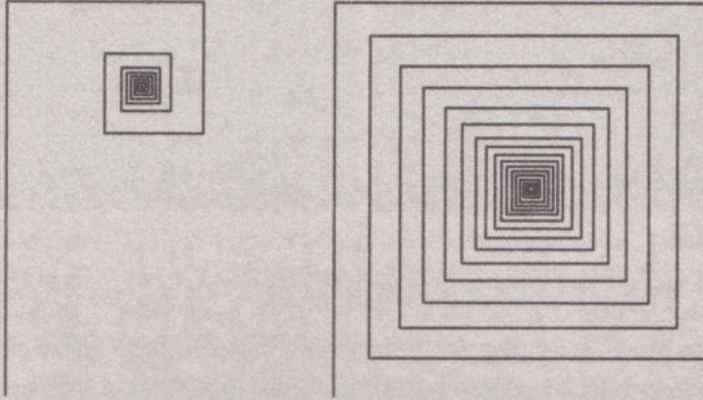
Let us now turn to the strange case of two polygonal spirals such as the ones in the image above. To the right is an infinite polygon where each side is smaller than the previous one by a proportion of $1/q$. The sum of all the sides will be the limit of the series $1 + 1/q + 1/q^2 + 1/q^3...$, which gives $q/(q-1)$. As such, the length of the polygon is finite. As an example, if we choose $q = 1.05$, the sum (or rather the length of all the sides) will be 21.

The spiral on the left has been created using a different rule that is also extremely simple: the longest side of the spiral measures 1; the next 1/2, the next 1/3; the next 1/4; and so on. It is known that the sum of this series does not have a limit, or rather, the spiral on the left is infinitely long, while the one on the right is not. Could we have guessed this?

pair of compasses and sets it to an width of one metre. They walk the distance of the coast, adding up the total as they fix their compasses on contiguous points on the coastline. The resulting value in metres is only an approximation of the genuine length, since the measurement does not include the concave and convex elements that measure less than one metre. They then set their compasses to a smaller aperture, for example a decimetre, and repeat the procedure. It is obvious that this time they will obtain a greater measurement since the truncated straight line (also referred to as polygonal) traced by their compasses will better fit the geographic features. Using common sense, it could be assumed that these values converge on a finite number that represents the true length of the coastline or the border. However Richardson showed that the measures of length grew without limit as the compass unit became
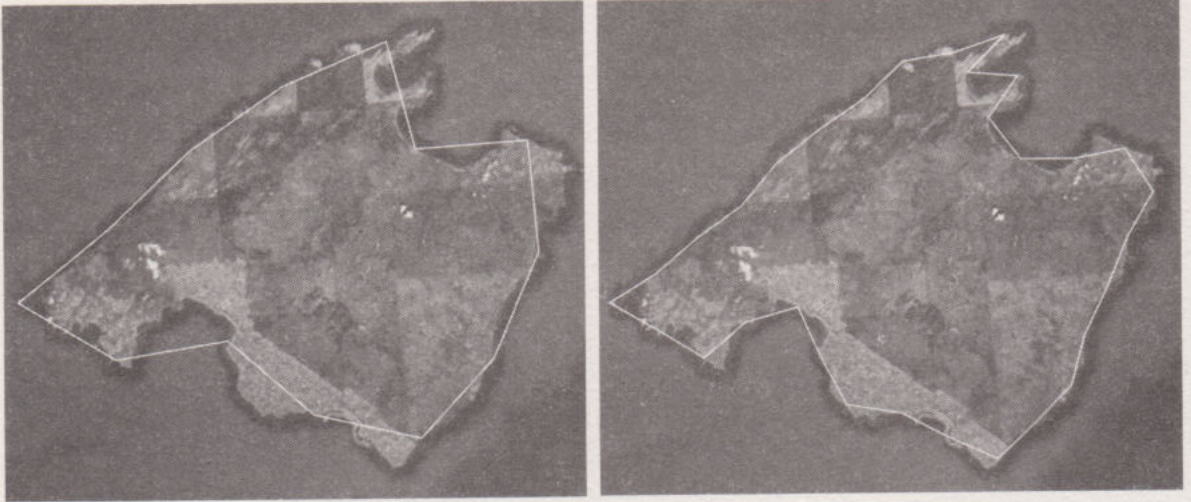
Two approximations for calculating the length of the coast of the island of Majorca made using segments of a different length. The one on the left has been made using a longer segment than the one on the right. It can be observed that the latter offers a better representation of the real coast than the first. Although it is surprising, the Richardson effect means that the limit of the measurements made with increasingly greater precision is not the real value of the coast, but infinity.

smaller and the scale of the map greater. This extraordinary fact is known as the 'Richardson effect'.

In his time, Richardson's research was ignored by the scientific community, however it is now regarded as being of vital importance for the significant role it played in the modern study of fractals. Benoît Mandelbrot cited Richardson's work in a well-known article published in 1967 entitled *How Long is the Coast of Britain?* In the article, Mandelbrot explained how the notion of length lacked meaning in the case of objects as irregular as borders. Consequently, mathematicians have devised a number to quantify how an irregular object with these properties fills space. This number is an extrapolation of the 'normal' dimensions of classical geometry (zero, one, two…). As a result, the dimensions of this class of irregular, 'non-Euclidean' objects are often fractions.

Everything depends on the scale used for measuring

What is happening here is that Euclidean geometry, with its whole number dimensions, does not manage to capture the essence of irregular forms. Richardson's experiment equates to calculating the length using different distances or scales. If someone calculates the length of a coastline from space, the result they obtain will

be smaller than that obtained by someone who combs the beaches like an ant to measure each grain of sand.

Now consider a ball of wool. When viewed from afar, it looks like a point, that is to say a zero-dimensional shape. However, as the observer draws close, they will see that the ball occupies a space similar to that of a sphere, a three-dimensional space. If they continue to move closer, they will see the strand that makes up the ball from close up, converting it into a one-dimensional object. Drawing closer still, such that the observer shrinks in order to appreciate the structure of the strand, they will start to see it once again in three dimensions as they make out the individual fibres of which it is made up, adopting the appearance of three-dimensional columns. And so on. It is clear that it is not possible to speak of the dimension of the strand objectively, since everything depends on the perspective of the viewer – the scale on which it is observed.

Let us now consider an example of the Richardson effect comparing the approximations to the perimeter of a circle on the one hand, and the real case of the island of Majorca on the other. Given a circumference that is 100 km in diameter, an order of magnitude similar to that of Majorca, we have a perimeter of π times the diameter, or 314.15… km. Let us plot the results on a logarithmic graph to better appreciate the range of different compass openings we shall use. Hence on the horizontal axis we shall mark the inverse of the opening, which can be interpreted as the precision of the measurement. When the opening s is small, that is to say the measurement is more accurate, $1/s$ is large. The vertical axis will be used for the logarithm of the length of the perimeter.

If we consider the case of a circumference, for compasses set open to 50 km, the best approximation to the perimeter of the circumference is a polygon with six sides of 50 km each, giving a total of 300 km; if the polygon now has 12 sides of 25.882 km, the approximation becomes 310.584 km; with a polygon of 24 sides of 13.053 it will be 313.272 km; with 48 sides of 6.54 km, we have 313.92; with 96 sides of 3.727 km, we have 314.112 km, and with 192 sides of 1.636 km, we also have 314.112 km. Hence as we reduce the size of the compass, our approximation to the perimeter of the circumference converges on the real value.

However this is not the case with the coastline of Majorca. Approximating it to a polygon whose sides are 28 km, we get a perimeter of 362.2 km; if the polygon is drawn with sides of 14 km, the perimeter is 416.7 km; with sides of 7 km it becomes 467.7 km and with sides of 3.5 km, the perimeter is 524.8 km.
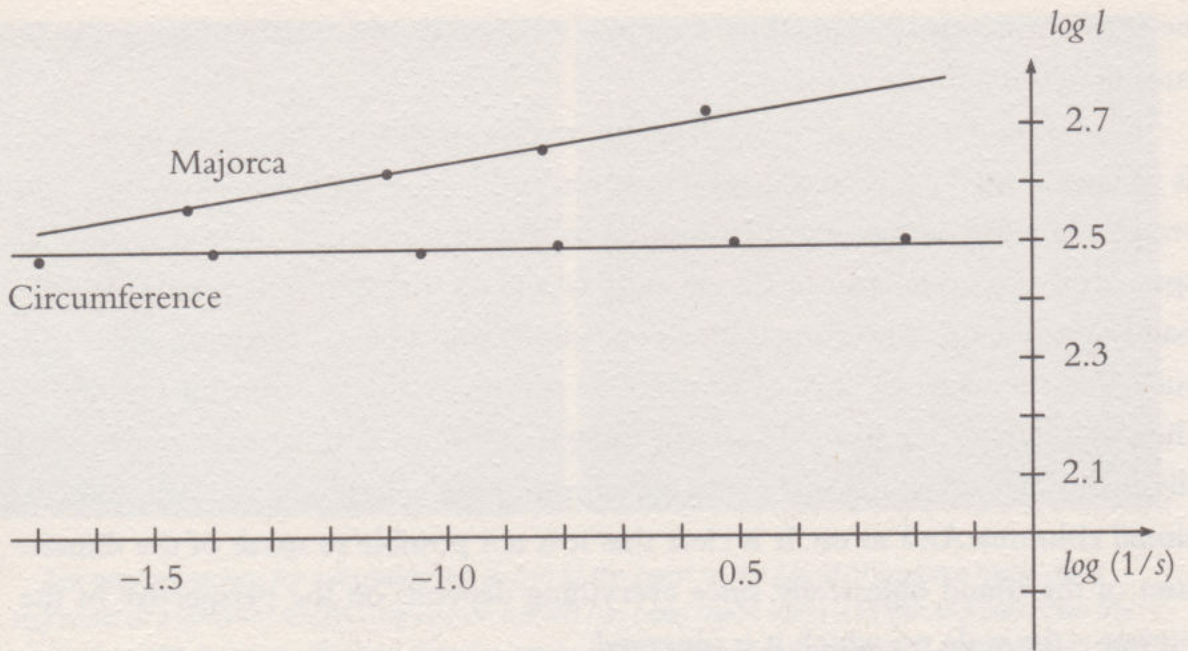
The points of the graph are a close fit to a straight line in both cases. Obviously we cannot expect the points to fall exactly on the line due to the nature of the measurements. In the case of the circle, the line is almost horizontal, whereas in the case of the Majorcan coast, it has a slight gradient of $d = 0.17$. The equation of the straight line can be expressed in the following way: $\log l = d \cdot \log(1/s) + k$, where l is the length of the perimeter estimated with an opening of s; k is a specific constant, and d, the gradient of the line that has been calculated. Converting the expression into exponential form gives:

$$l = c / s^d.$$

Note the similarity of this formula to Korcak's law. Richardson's work concludes that the traditional concept of the length of a coast is meaningless and proposes the incorporation of the new concept of 'roughness', which is determined by the value of the gradient d that we have calculated in the previous example. In the case of real coasts and borders, he estimated the following values of d:

$d = 0.25$ for the west coast of Britain, one of the most ragged on the planet.

$d = 0.15$ for the German border;

$d = 0.14$ for the border between Portugal and Spain;

$d = 0.13$ for the Australian coast; and

$d = 0.02$ for the South African coast, one of the most uniform.

In nature, fractal objects often appear in relation to two concrete circumstances or situations:

Borders. This includes all the cases in which two media — human, natural, physical, chemical, etc — come into contact, or rather two different surfaces. A border between counties, river banks, coastlines, clouds…

Trees. Includes cases where there is a branching with self-similarity: trees, bushes and plants, river basins…
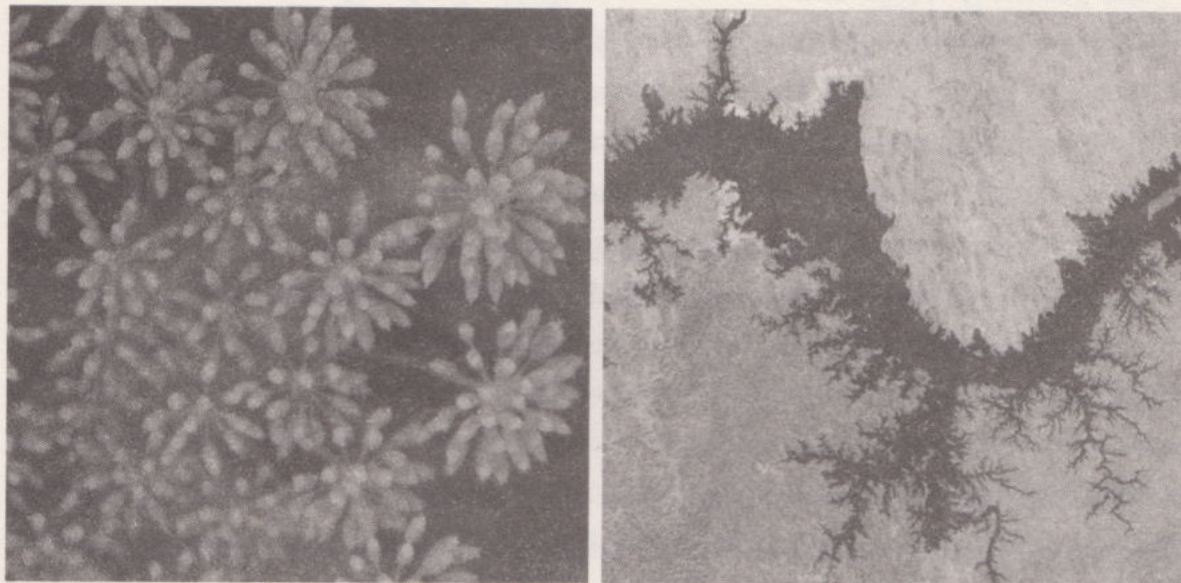
Certain plant structures or aerial views of basins of some rivers exhibit a fractal structure.

Coverings

Curves, surfaces and volumes can often be so complex that their measurement becomes an obsession. However, the fundamental idea that quantities such as length, area or volume do not vary arbitrarily with respect to the scale and are related by a law allows us to calculate one based on the other. The idea discovered by Richardson (together with those discovered by Korcak, Zipf and Hurst), according to which the length is a power function of the precision with exponent d, will be useful in allowing us to consider a new concept: dimension.

At the start of the 20th century, one of the greatest mathematical problems was determining the meaning of dimension and its properties. The situation grew worse with the appearance of different types of dimensions — topological, Hausdorff, fractal, self-similar… and many others. All are related and some make sense in some situations while in others they do not. Similarly, on some occasions they are equal, on others they are not. Contrary to what we might hope, we should not

expect a single definition of dimension to reveal the true essence of the concept. The quest for the Holy Grail, or rather the single, universal and acceptable definition, is a vain one.

In his book *Measure, Topology and Fractal Geometry*, Gerald A. Edgar illustrates the concept of dimension in the following terms:

"If we have a point that we want to imprison, we can use a small cube as the prison. [...] The cube consists of six plane faces; we need to know that they are two dimensional. A point living in these faces (Flatland) can be imprisoned using a small circle. So saying that the faces of the cube are two dimensional requires knowing that a circle is one dimensional. A point living in the circle (Lineland) can be imprisoned using just two points as prison walls. So we need to know that a two-point set is zero dimensional. Finally, a point living in the two-point set (Pointland) is already unable to move. So we need no prison walls at all. This will be the definition of a zero-dimensional set."

The idea of an inductive definition of dimension is overcome in Euclid's *Elements* where the concept is defined implicitly. It is said that a figure is one dimensional if its border is made up of points; two dimensional if it is made up of curves; and three dimensional if it is made up of surfaces.

Poincaré re-stated the matter in similar terms with the name 'topological dimension'. His formulation was as follows: A space has dimension n if it can be divided in some way by another with dimension $n-1$. To make this definition more rigorous, it is necessary to correctly define "divided in some way". In 1913, Brouwer made the first attempt at specifying this action, followed by Pável Urysohn ten years later. Each has a different interpretation, however they coincide for locally connected spaces. Hence there are currently three definitions of topological dimension that are regarded as significant. The inductive one by Urysohn (and Menger), the inductive one by Brouwer (and Cech), and the Lebesgue covering dimension[3].

---

3 According to Brouwer's definition only the empty set has dimension −1. The dimension of a space is the smallest whole number n such that any element has a series of arbitrary small openings with borders whose dimension is strictly less than n.

The construction of the topological Lebesgue covering dimension (hereafter referred to as the topological dimension or covering dimension) is particularly suited to highly irregular sets. The covering dimension is easy to visualise. A cover of a subset S of $\mathbb{R}^n$ is any collection of open sets[4] the union of which contains the set S. The illustration shows one way of covering a curve in $\mathbb{R}^2$.
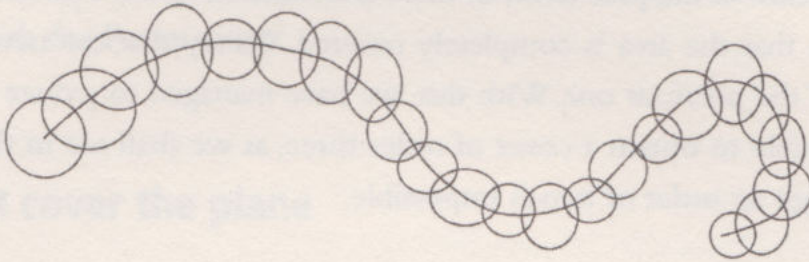
A curve with a second order covering dimension.
(Source of illustrations on this page: Maria Isabel Binimelis.)

We can do the same for an area of the specific plane. Let us use an easy-to-understand analogy. We have to paint a certain area the colour green and we have one or more pads that may or may not be circular. A cover of this area would be to paint it green without leaving any gaps. Clearly, some sections are covered several times, meaning that they will be darker if we have used a paint that is not completely opaque. Of all these we select the darkest one (or ones), that is to say the one which has been painted the most times, and the number of times the darkest section (or sections) has been covered is referred to as the order of the cover. For example, in the case of an area, we have the following covers:
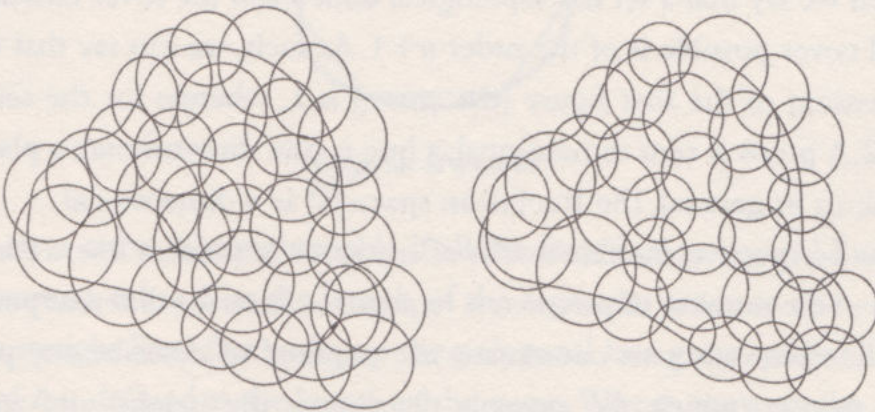
4 To define 'opening' we first need to determine in which space we are working and the metric that is being used. In this case, for the sake of brevity, let us consider real space with the Euclidean metric.

Now carefully observe the first cover (left) with the ovoid area, specifically focusing on the area shaded black. This is covered by five pads and there is no other area that is covered by a greater number. As such, the order of this cover is five. Can this be reduced? That is to say, can we use our pads to mark each and every point on the surface, such that we do not need to mark any of these at least five times? In the image on the right, we can see that the answer is yes: it has been possible to slightly shrink the surface of the pads (each of these is contained in the equivalent figure on the left), such that the area is completely covered. We say that the new cover is an over-cover of the previous one. With this, we have managed to reduce the order to four. It is possible to obtain a cover of order three, as we shall see in the following figure, although an order of two is impossible.
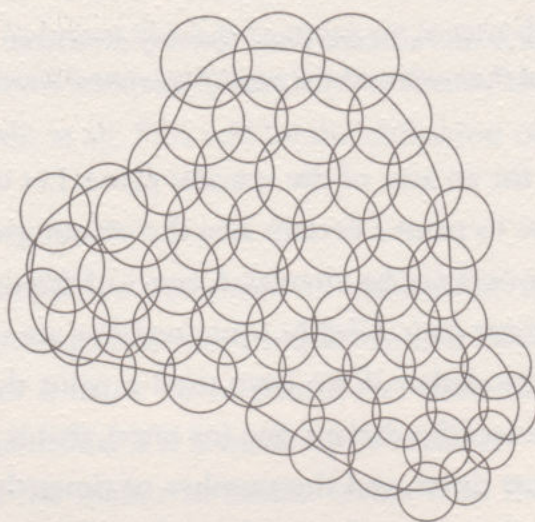
Our ovoid area covered with a maximum of three pads.
(Source: Maria Isabel Binimelis.)

In general we say that a set has topological dimension (or cover dimension) n if the minimal cover possible is of the order $n + 1$. As such, we can say that the topological dimension of the first figure (the curve) is 1, whereas for the second (the ovoid) it is 2. A point is zero dimensional; a line is one dimensional; a plane is two dimensional; an in general, the Euclidean space $\mathbb{R}^n$ is n dimensional.

From this perspective, the dimension of any space (a point, a line, a surface, etc.) corresponds to the number of parameters required to describe different points in the space. For example, only two coordinates are required to describe any point on a plane – an abscissa (which, for example determines the height) and an ordinate (which determines the width). In contrast, for space, three coordinates are required: the width, the height and the depth.

The definition of the topological dimension was largely motivated by two strong blows to the traditional interpretation of dimension (excessively dependent on intuitive and imprecise notions, such as 'thinness') received at the end of the 19th century. The first came from Cantor's proof that there is a point-for-point correspondence between the real straight line $\mathbb{R}^1$ and the real plane $\mathbb{R}^2$; the second from Peano's construction of a continuous function in $\mathbb{R}^1$ on $\mathbb{R}^2$. We are going to learn more about each of these important discoveries in a simple and intuitive manner, starting with the Peano curve.

Curves that cover the plane

One of the goals of calculation is to measure things – lengths of curves, areas of regions, volumes of solids… Under certain circumstances it can be difficult to measure the length of a curve. However if it is approximated by means of a line composed of straight lines (also known as a polygonal line or simply a polygonal) it is possible to obtain a close approximation. The smaller the straight lines (or segments) of the polygonal line, the closer the approximation. In the following image, we can see an example of this process. We are approximating the length of a sinuous curve using various straight lines connected to each other, such that their ends all lie on points of the curve.
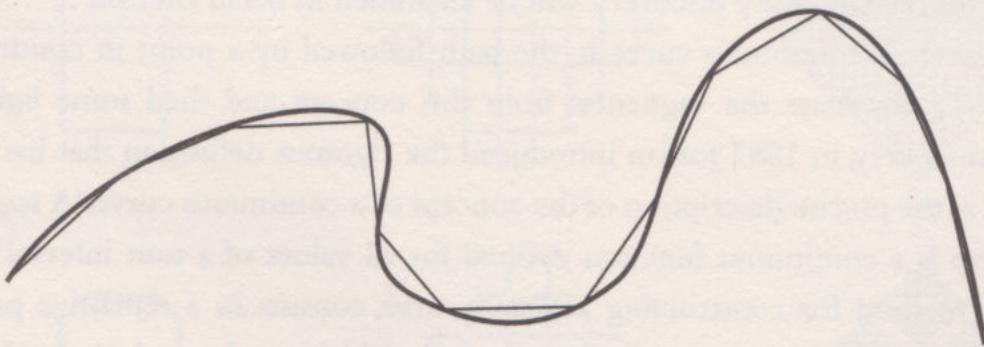
Polygonal of a curve.

We will say a curve is straightenable if all the inscribed polygonal lines tend to a certain common value L as the length of the segments tends to zero – in other words as the straight lines that make up the polygonal line grow smaller. This common value of L is defined as the length of the curve. We can use a similar reasoning for areas, in this instance approximating the plane using small rectangles instead of straight lines.

In this example we are making use of various objects with a topological dimension of one (straight line segments) to approximate another object in the same dimension (a curve). The operation has been carried out using an ingenious and even surprising procedure, which is also intuitive.

But what happens if we now consider the possibility of approximating objects from any Euclidean dimension using others of a lower dimension? For example, is it possible to approximate the area of a square using a curve? The intuitive reaction when faced with such an idea is to say no: Curves do not have thickness and as such, it is not possible to use them to construct a thick space. Or put another way, an object with topological dimension one (a curve) cannot be transformed into an object with dimension two (e.g. a square). To claim the contrary would be, intuitively speaking, a genuine monstrosity.

The Peano curve

In 1890, the Italian mathematician Giuseppe Peano discovered a continuous curve which passed through all the points of a square the sides of which were length 1; in other words, a one-dimensional object could be transformed into another with dimension two. Peano followed the steps laid out by Cantor, who had previously and completely counter-intuitively proved that the cardinality of the infinite points of the unit interval was the same as the infinite points of any surface such as the unit square. This revolutionary discovery will be examined in detail later on[5].

Intuitively, a continuous curve is 'the path followed by a point in continuous motion'. To eliminate the vagueness from this concept and shed some light on Peano's discovery, in 1887 Jordan introduced the rigorous definition that has been adopted as the precise description of the concept of a continuous curve: 'A segment of a curve is a continuous function defined for all values of a unit interval'. The standard method for constructing a Peano curve consists in a repetitive process starting with an initial curve with nine segments, which are then substituted by the figure generated in each previous iteration.

---

5 The Cantor transformation, while being bijective, is not continuous. The goal of Peano, on the other hand is to construct a *continuous* function of the unit interval over any unit square, even if it is not bijective (in fact various points of the unit segment have the same image). That is to say that in the end, the unit segment and the unit square are not equivalent. It was necessary to await the contribution of the Dutchman Jan Brouwer in 1911 for a rigorous proof.
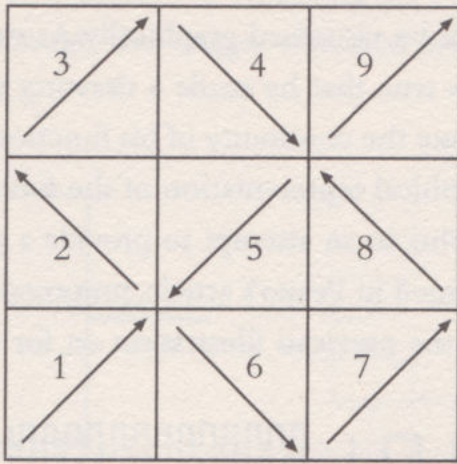
In order to draw the nine starting segments, the following method should be followed, first drawing the line indicated by number one and so on: the process is then repeated for each of the nine starting segments (i.e. substitute each of the nine starting segments for the drawing we have made), and the process is then iterated repeatedly. The result is a curve that looks like the following. (In the second set of three images, the corners have been truncated in order to better appreciate that the Peano curve can be constructed as a continuous path.
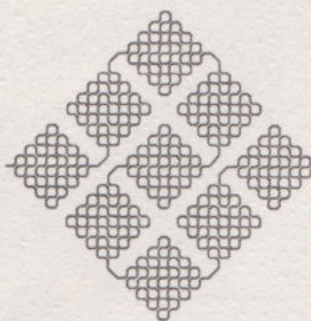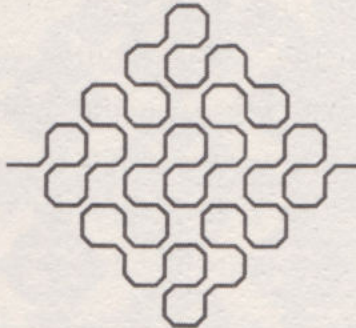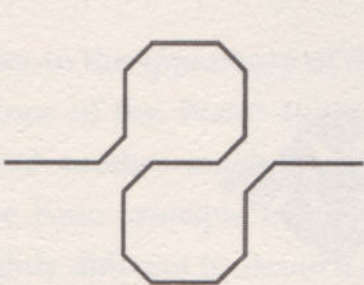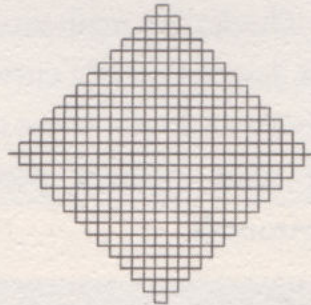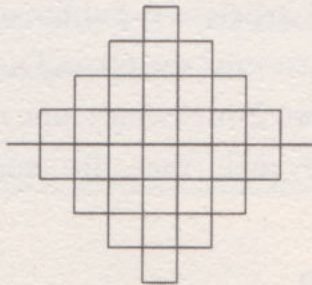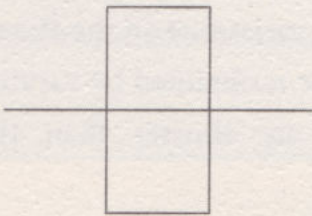
After an infinite number of iterations, the Peano curve would not take up any more space than the initial square.

73

However Peano's original contribution was completely analytical. It did not define an iterative process and nor could it be visualised graphically. As such there was no curve to be drawn. (However it is true that he made a drawing of an inclined figure of eight as a step to demonstrate the continuity of his function). Peano only indicated the order in which the graphical representation of the function fills the square. It was other mathematicians who, in an attempt to provide a graphical representation of the abstract function defined in Peano's article, proposed iterative series of curves like the one indicated in the previous illustration or, for example, this one:
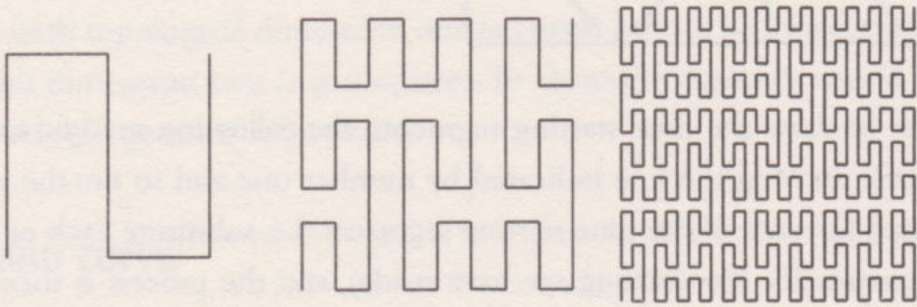
Consequently, we do not know exactly which of these two curves can be properly referred to as the Peano curve. The two polygonal approximations tend to the same limit: the square.

The Italian mathematician's article was published in 1890 and constituted the first description of a curve that covered the plane. There are variations on the Peano curve that do not cover the plane. For example, one of these is obtained by varying the nine segment procedure such that the verticals are shorter than the horizontals.
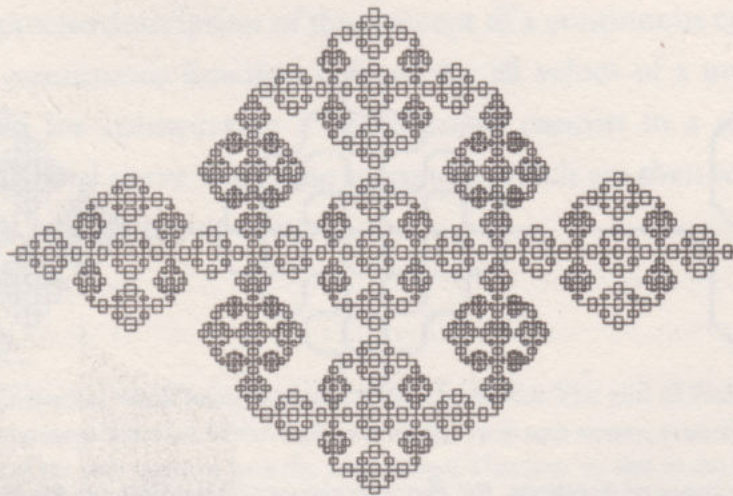
Others are obtained by removing the central segment. This last curve is peculiar because although the images obtained are continuous, the function that defines them is not.
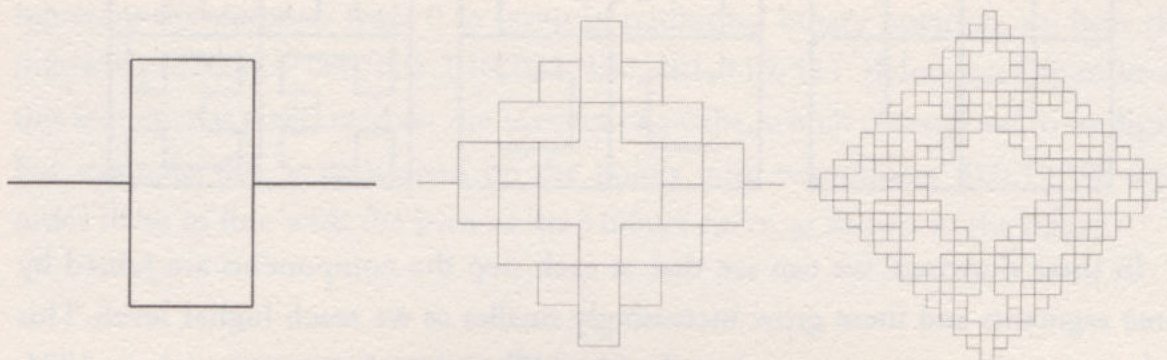
MUSIC AND MATHEMATICS

Musical compositions have been created based on the idea of a one-dimensional object. For example, the violinist Scott St. John composed an 11-minute piece for the double bass and English horn. In the first part the two instruments tended to fill their respective rhythmical and tonal spaces. When a tonality began to emerge, one of the two instruments would immediately escape to another tonality. The result is a confrontation between long expressive phrases and quick bursts of energy. In the second part, the two musicians cooperate in patterns of form and style. The imprecise tonalities of the first section are set, and if the score is examined at different scales, there are great similarities.

The Hilbert curve

Prior to the appearance of the graphical representations of the Peano curve, David Hilbert published another example of a space-filling curve. The basic principle behind the Hilbert curve is slightly different to Peano's: instead of there being only one motif, various are used, each with different rules. Such constructions are referred to as 'non standard'.

David Hilbert.

75
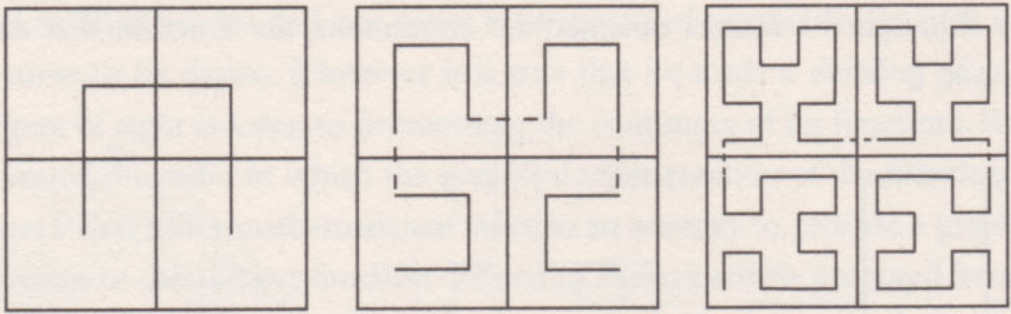
In these drawings, we can see that at each step the components are joined by three segments and these grow increasingly smaller as we reach higher levels. This is the original construction as described by Hilbert in a two-page article in 1891. There is a standard construction of the same curve starting with a slightly different shape, but we shall leave this to the more inquiring reader. The difference between Hilbert and Peano curves is that the former applies intervals of length that are half of the previous ones to squares with sides that are half the previous, whereas the latter does its own thing with intervals and squares of one third.

There are extremely interesting variations of the Hilbert curve, such as the one which uses an inverted V as its initial base, and another one by Carl Hansen that is based on an H (clearly in honour of the person who inspired its creation). Another attractive attribute of the Hilbert curve is that it can be modified to cover a volumetric space, as shown in the image below.
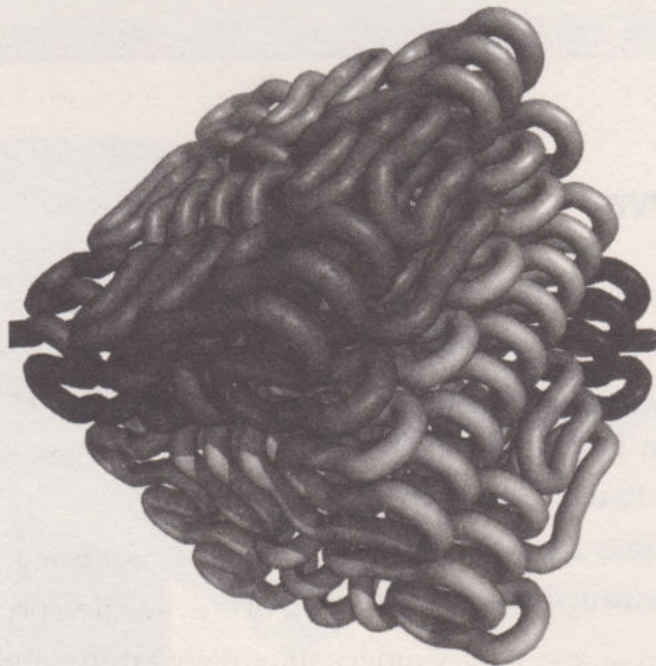
This three-dimensional version has extremely important applications in mechanical devices, especially those in which transmission errors depend on the so-called Gray code, which is based on a variation of the binary notation. If we represent the numbers from 0 to seven in traditional binary notation, we have the following sequence: 000, 001, 010, 011, 100, 101, 110, 111. Below, each number of this sequence is positioned on the vertices of a cube, which correspond to its digits. For example, 001 is positioned on the points with coordinates (0,0,1). We then order these in line with the path of the Hilbert curve, as shown in the figure.
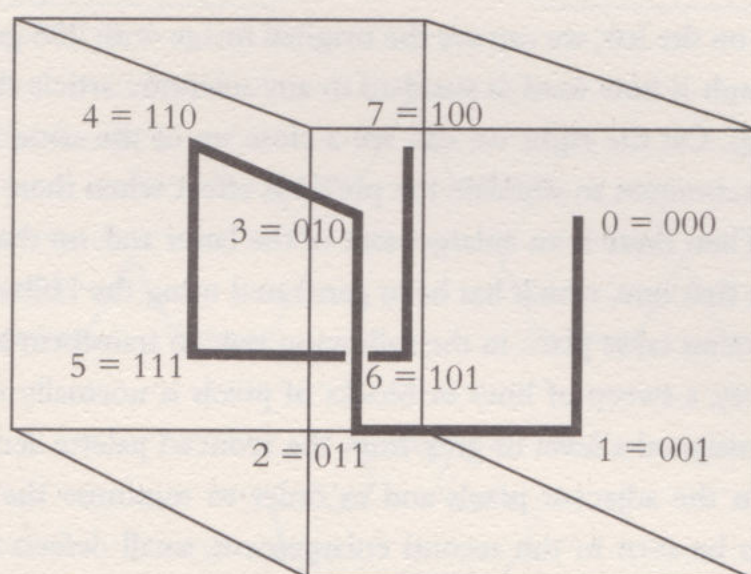
This sequence is mentioned in the Gray code and has the peculiarity that the binary sequence that is obtained follows a special order. If we observe carefully, we can see that each triplet of numbers varies from the previous one by only one digit (one bit of information), something that does not happen in the traditional sequence, in which for example, 001 is followed by 010 (a two-digit difference). Technically speaking, the Hamming difference between one triplet and the next is one. If, instead of coding the first 8 numbers, we need more, we can select the next iteration of the Hilbert curve such that we have the numbers from 0 to 31. The Gray codes enables a significant reduction in transmission errors and are widely used in digital television broadcasting, for example.

Another application of the Hilbert curve is the processing of digital images. When we wish to send a grey-scale image to a first-generation printer, it is necessary to construct an approximate binary model, since these printers only understand one code: 0 or 1 (ink or no ink). In order to achieve this, techniques referred to as 'dith-

ering' are used. The aim of these techniques is to simulate the use of an extensive colour palette when, in reality, only limited colours are available. They also seek to replicate many different shades of grey when only a binary description is available.
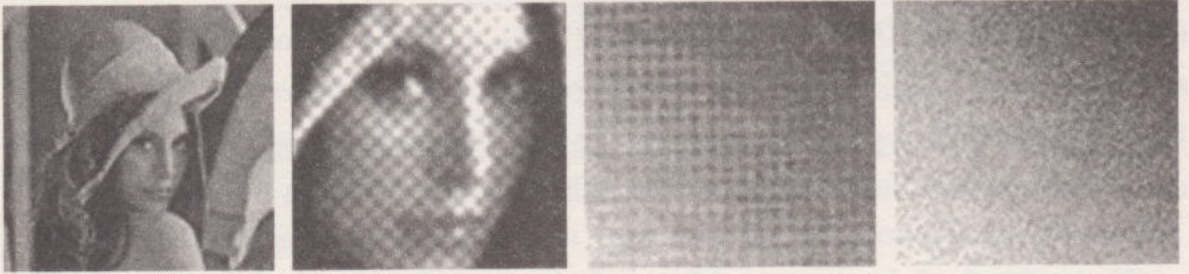
In the image on the left, we can see the original image with 256 grey levels (this specific photograph is now used as standard in any scientific article that deals with image processing). On the right we can see a close up of the same image treated with dithering techniques to simulate the previous effect when there is a set of less than 256 greys. Then there is an enlargement of the latter and, on the right, an enlargement of the first one, which has been generated using the Hilbert curve.

The latter process takes place in the following way: to transform an image with 256 shades of grey, a sweep of lines or blocks of pixels is normally carried out. A specific pixel is assigned a level of grey from the reduced palette according to the shades of grey in the adjacent pixels and in order to minimise the overall error. Normally, as can be seen in the second enlargement, small defects appear in the image, when it is calculated in this way, thus making the dithering used conspicuous. To eliminate such defects, instead of sweeping using horizontal lines, a Hilbert curve is used, which passes through all the pixels of the image. The advantage of this method of sweeping is an absence of the directional trends inherent to the other procedures.

Triangles, sponges and snowflakes: the fractal dimension

Following Peano and Hilbert's publications, many other mathematicians proposed new examples, including Koch, Levy and Cesaro, Sierpinski and Schoenberg. In 1904, the Swiss Helge von Koch published the article "On a continuous curve without tangents, constructible from elementary geometry". This somewhat fearsome title hides a concept, which is as simple as it is surprising. Let us consider a horizontal straight line segment with length one. Replace the initial segment with

four straight line segments each with a length of 1/3, thus forming the first curve of an iterative series as shown in the diagram below:
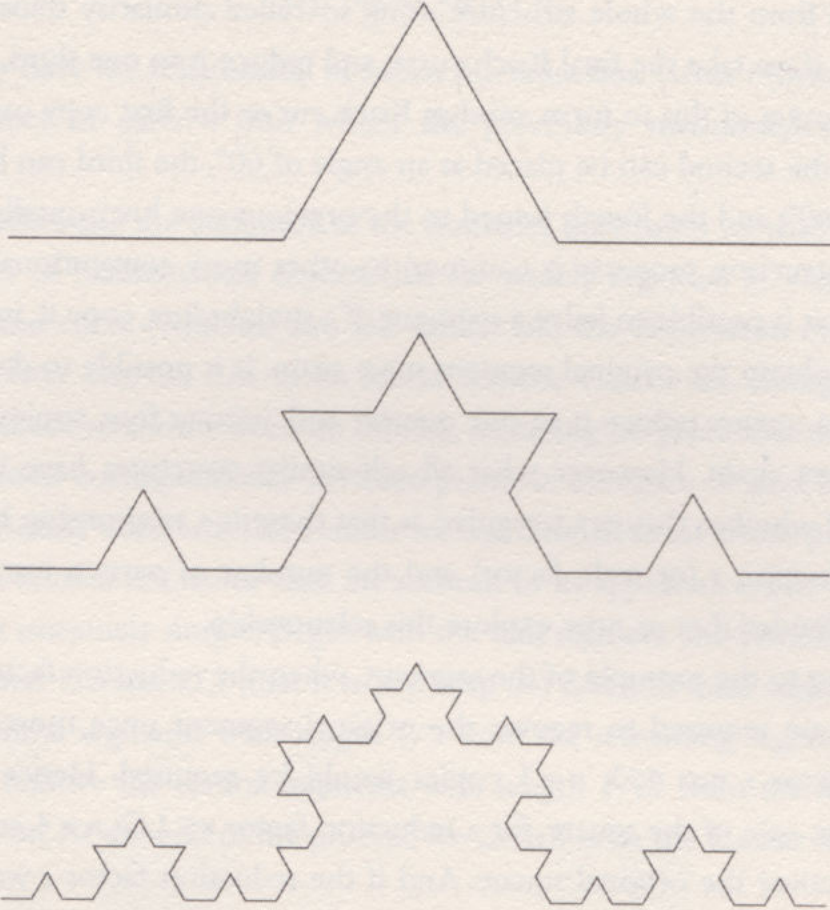
If we arrange one of these curves on each of the three sides of an equilateral triangle with length one, we obtain the so-called Koch snowflake. This curve has the amazing property that its length is infinite although the area contained within it is not.

After defining the concept of the cover dimension for any type of object and having observed curves such as Peano's and Hilbert's in which each sequence has a topological dimension of one but that nonetheless cover a region with dimension two, mathematicians found themselves obliged to define another type of dimension to fit with these observations. Koch's curve is particularly suited to providing an example of just what we are talking about. This new concept of dimension, which will be explained below, is referred to as the fractal dimension of a self-similar object. Further

on, extending this to figures that are strictly self-similar, we will obtain what is referred to as the fractal dimension.

Although the concept of similarity will be discussed in greater detail in the following chapter, let us here state some of its essential principles. A structure is said to be self-similar if it can be divided into small replications of itself. These parts are in fact obtained from the whole structure using so-called similarity transformations. For example, if we take the final Koch curve and reduce it to one third, we can use three more copies of this to form another Koch curve: the first copy can be placed horizontally, the second can be placed at an angle of 60°, the third can be placed at an angle of -60°, and the fourth joined to the previous one horizontally.

This construction property is common to other more conventional structures. For example, it is possible to halve a segment of a straight line, copy it, join the copies and then obtain the original segment once again. It is possible to do something similar with a square; reduce it to one quarter and, joining four copies, obtain the original square again. However what all self-similar structures have in common (regardless of whether they are irregular) is that there is a relationship between the factor of reduction r (or scale factor) and the number of parts n into which the structure is divided. Let us now explore this relationship.

Returning to the example of the segment, when the reduction factor is $r = 1/2$, $n = 2$ copies are required to recover the original segment once more. And if the reduction factor r was $1/3$, $n = 3$ copies would be required. Hence in general, $n = 1/r$. In the case of the square, for a reduction factor $r = 1/2$, $n = 4$ copies are required to recover the original square. And if the reduction factor r was $1/3$, $n = 9$ copies are required. This gives us $n = 1/r^2$.

If we do the same for a cube, with a reduction factor of $1/2$, 8 copies are required to recover the original size and with a factor of $1/3$, 27 copies are required. In both cases, we have $n = 1/r^3$. The exponent of r always coincides with the topological dimension of the original figure.

However, if we re-calculate for the Koch curve, in the first iteration we have $n = 4$ and $r = 1/3$. In this case, the relationship is not so obvious. However, if we let ourselves be guided by the cases of the segment and the square, we can postulate that the power law must also hold for the structure of the Koch curve, such that $4 = 3^D$, where D is a provisional measurement of the dimension of the curve in question. A simple way to calculate D is to take logarithms of both sides of the equation, such that $log\,4 = D \cdot log\,3$, or also that $D = log\,4\,/\,log\,3 = 1.2629$. If we do likewise for the second iteration of the curve, we have $16 = 9^D$, or rather, $D = log$

$4^2 / log\, 3^2 = 2 \cdot log\, 4 / (2 \cdot log\, 3) = 1.2629$. Generally speaking, the law that has been established will hold for any iteration, such that we always have $D = 1.2629$. This number is called the similarity dimension and is referred to as D_S so as not to be confused with other fractal dimensions, we can define it as:

$$D_S = log\, n / log\, (1 / r).$$

We already have the relationship between the reduction factor r (or scale factor) and the number of parts n into which the previously mentioned structure is divided.

To fully understand the definition of the similarity dimension we shall now apply it to a number of mathematical objects that are widely regarded as 'classics'. Long before the Peano curve (which is also self-similar and the dimensions of which we shall calculate later on), the first to be found was the Cantor set. Georg Cantor is today known above all for his work on infinity, including his proof that there is a re-ciprocally univocal correspondence between points in the space $\mathbb{R}^1$ and those in $\mathbb{R}^2$, as we had previously observed. The Cantor set was defined in an article published in 1883. It is also known as Cantor dust on account of its appearance. The rule for its construction is extremely simple: Begin with the unit segment and remove the open segment between 1/3 and 2/3 (that is to say, keep the first and third segments); now remove the central segments with length 1/9 from the remaining segments; in the third step, we remove the central segments with length 1/27 from the four which remain, and so on. At the end of the process, we are left with the Cantor set:
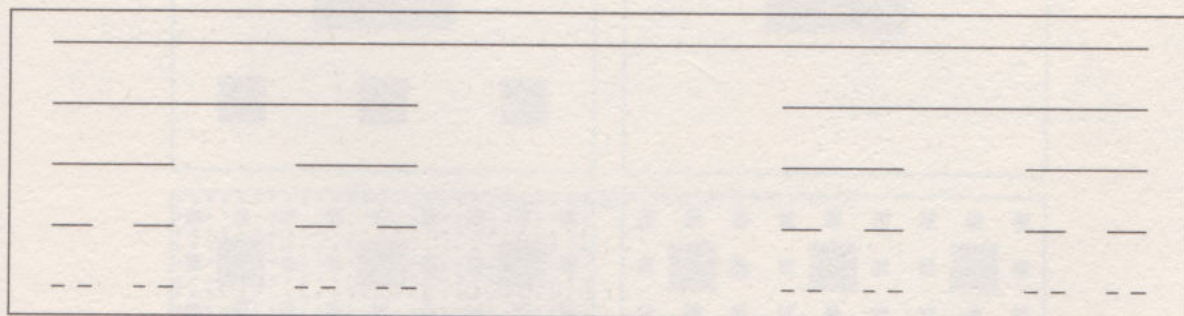
The set is extremely difficult to visualise given its tendency to 'disappear', but we can get an idea of its appearance if we follow the iterative process used in its construction. Observe that if we reduce the complete Cantor set to 1/3, we obtain exactly the first section of the left, and if we make a copy of this reduction and shift it 2/3 to the right, we have precisely the right section. That is to say that the Cantor

set is made up of two sections that are proportionally smaller by a third part similar to the total. If we apply the formula obtained for the similarity dimension, we have:

$$D_S = \log 2 / \log 3 \simeq 0.6309.$$

In the Cantor set, there is no connection between the points, and as such it has a topological dimension of zero. Its similarity dimension, as can be observed, is greater than its topological dimension.

For the Peano curve, according to the nine segment construction, we have $n = 9$ and a reduction factor of $1/3$; thus the similarity dimension is:

$$D_S = \log 3^2 / \log 3 = 2.$$

Taking two Cantor sets, one horizontal and the other vertical, and multiplying them together, we obtain a structure that is known as the Sierpinsky carpet. It was first described by the Polish mathematician Waclaw Sierpinsky in 1916. The four initial iterations of the Sierpinsky carpet are as follows:
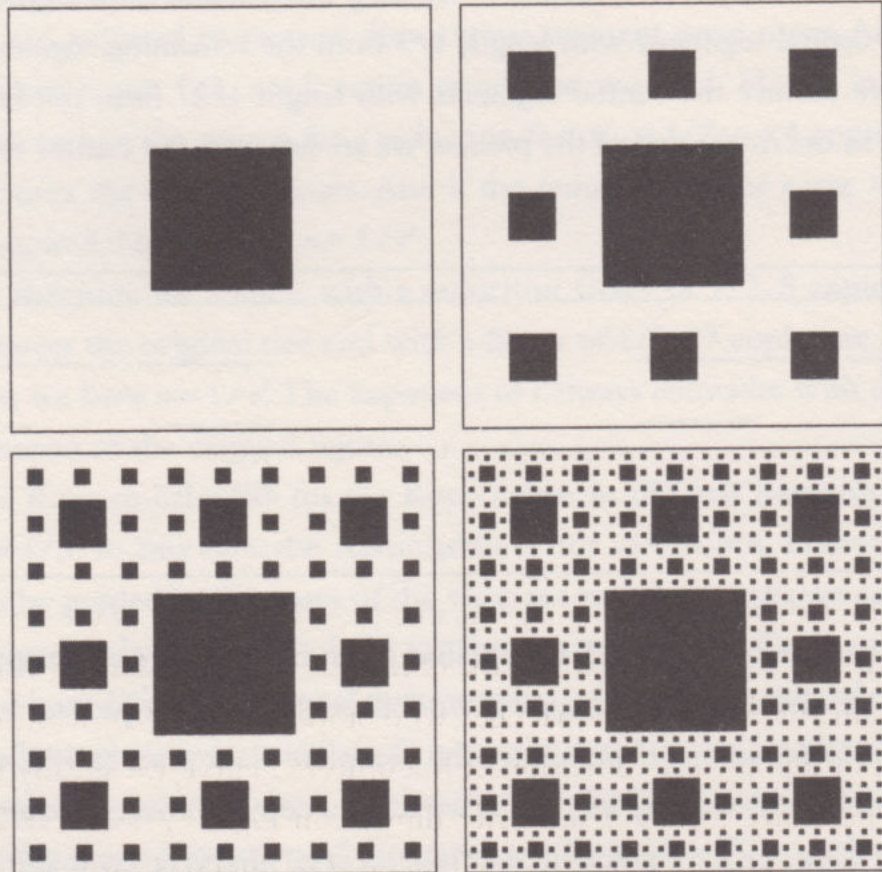
We can equate the construction of the carpet to an iterative process that removes the central square from a set of nine. An alternative method for obtaining the Sierpinsky carpet is to remove the central segment in the construction of the 9-segment Peano curve. Since it can be obtained by repeating the same pattern 8 times, reduced by 1/3, its similarity dimension will be $log\,8\,/\,log\,3 \simeq 1.8928$. Sierpinsky showed that the object thus obtained is what is known as a universal curve, that is to say one that contains any curve that can be constructed on the plane. We can obtain an infinite number of 'carpets' by devising a similar process using pentagons or any other regular polygons. The best known, based on a triangle, is referred to as the Sierpinsky triangle and was studied by the eponymous mathematician shortly before the four-sided carpet. The triangle can also be obtained by the iteration of a curve, and has a topological dimension of one and a similarity dimension of $log\,3\,/\,log\,2 \simeq 1.5850$.
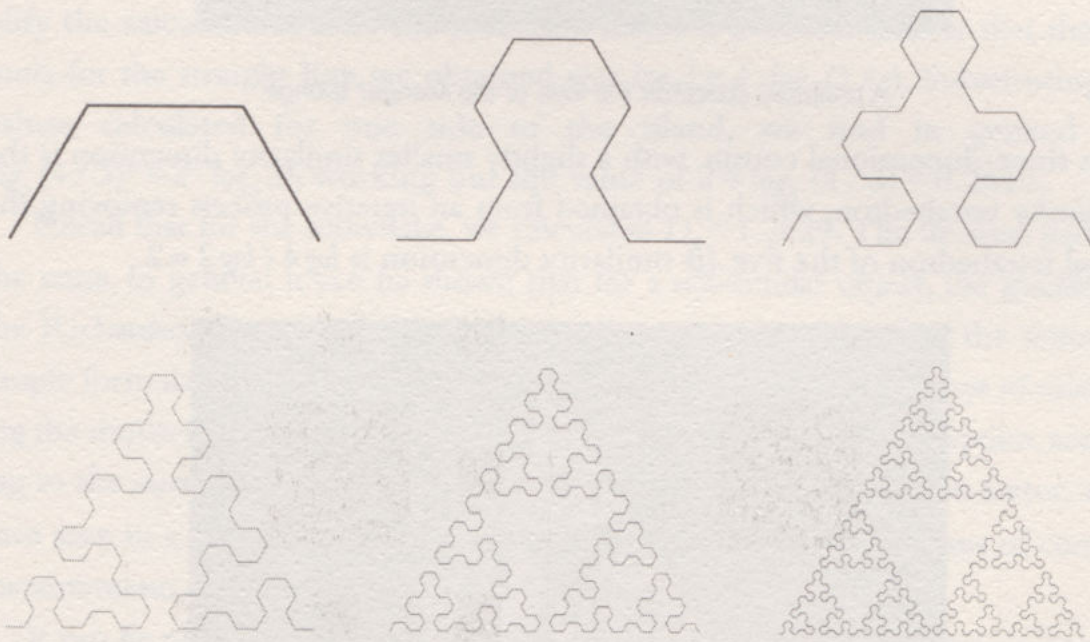
The first iterations of the Sierpinsky triangle.

If we jump to three dimensions and generalise the construction of the set of Cantor cube we discover another extraordinary object, the sponge Menger, named after the Austrian mathematician Karl Menger, who proposed it in 1926 while exploring the concept of topological dimension. It is a universal curve we already know, but in three dimensions. You have similarity dimension log 20 / log 3 $\simeq$ 2.7268, as obtained from 20 dice similar to the total reduced to the third party.
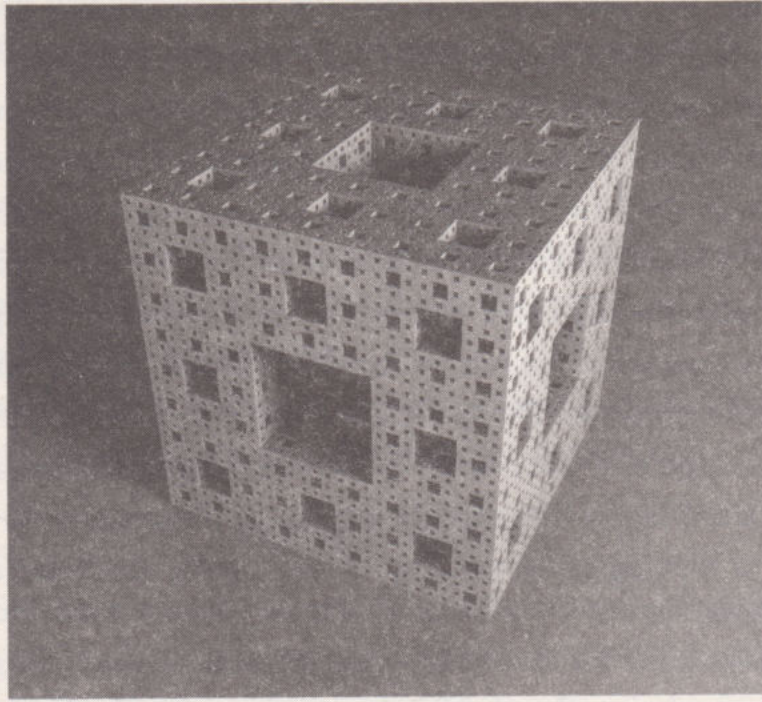
A sculpture based on the idea of the Menger sponge.

Its three-dimensional cousin, with a slightly smaller similarity dimension is the Sierpinsky tetrahedron, which is obtained from an iterative process removing the central tetrahedron of the five. Its similarity dimension is $log\, 4\, /\, log\, 2 = 2$.
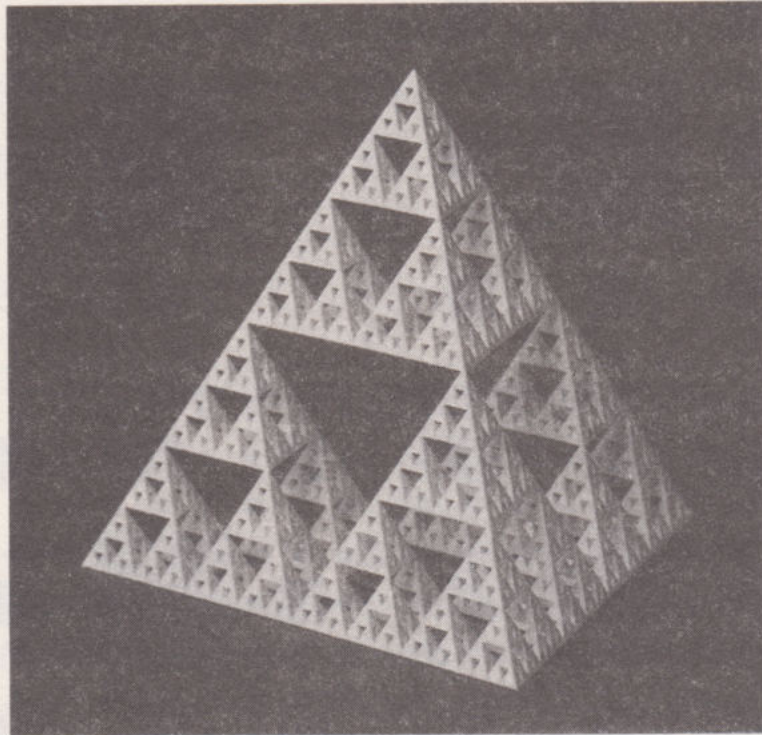
The Sierpinsky tetrahedron.

Now that we are familiar with the calculation of the similarity dimension, we are going to relate it to the exponent obtained using Richardson's law when attempting to measure a coastline or border. Let us imagine that we would like to calculate this law for the coastline of an imaginary new island, identical to the Koch snowflake. The coast of this island, let us call it Von Koch Island, is made up of three identical curves, each subdivided into self-similar parts with a reduction factor of 1/3. As such, it is natural to open the compasses to measurements of 1/3, 1/9, 1/27... Let us now measure one of the three sides of the island, starting with the compass set to 1/3. Supposing that the length of the side of the original triangle is one unit, we obtain 4/3 as our first approximate measurement of the coast. If we now repeat the procedure setting the compasses to 1/9, we obtain a measurement of 16/9. Proceeding in this way, for a compass opening of $s = 1/3^k$ we have $l = (4/3)^k$.

Now let us represent these measurements on a logarithmic graph. As we are free to select the base of the logarithm, we shall choose base 3 in order to simplify the calculations, since the reduction factor is 1/3. Remember that the formula for the straight line we obtained was $log_3 l = d \cdot log_3 (1/s)$. Substituting the values calculated for one side of the island, we find in general that $log_3 (4/3)^k = d \cdot log_3 3^k$, working out the value of $d = log_3 (4/3) = 0.2619$.

Recall that for the snowflake, we calculated $D_S = 1.2629$. The decimal parts are the same. In general it can be shown that for a self-similar object, the gradient of the Richardson line d and the similarity dimension are related by the following simple formula: $D_S = 1 + d$. This means that there are two different ways of calculating the similarity dimension: the first is based on the geometric description, according to the number of structures similar to the total and the reduction factor, as we have seen in a number of examples. The second is calculated by means of compass measurements.

It can be observed that the dimension calculated by the Richardson procedure generalises, in a certain sense, the self-similarity dimension by adding one. This makes it possible to calculate the fractal dimension for curves, that are not self-similar, such as coastlines or borders. However, how can we calculate this for objects that look more like a mark, or a real sponge, or a cloud? In these cases it is not possible to apply the compass procedure. Calculating the fractal dimension of an object can be a complicated task. In fact, there are many fractals for which it has still not been possible to calculate their dimension.

The definition we are searching for gives the Minkowski-Bouligand dimension, also known as the Minkowski dimension or the box-counting dimension.

On account of the ease with which it can be programmed on a computer, this is most commonly used in science. It is also in some senses similar to the topological covering dimension and the definition for self-similarity. Let us consider why. Consider the cover of an object the dimension of which we wish to calculate. If the object is on a plane, we can consider a cover using relatively small circles, and if it is in space, we can consider a cover using spheres. This strongly reminds us of the definition for covering. In order to be able to speak generally of the intervals of a straight line, discs in the plane and spheres in space, we speak of 'balls' with a radius of epsilon (ε). Let us denote the number of these balls with radius ε by $N(\varepsilon)$. Calculating the natural logarithm for this quantity and dividing it by $log\,(1/\varepsilon)$, leads us to the definition of the similarity dimension. Recall that applying the formula of the latter to the different factors of reduction always gives the same result. This is not the case for objects that are not self-similar, such as those in which we are now interested. Let us define the Minkowski dimension, D_M, as:

$$D_M = \lim_{\varepsilon \to 0} (log\,N(\varepsilon)\,/\,log\,(1/\varepsilon)).$$

THE EFFICIENT SALESPERSON

In 1912, shortly before discovering the triangle that bears his name, Sierpinsky studied a space-filling, recursive curve.
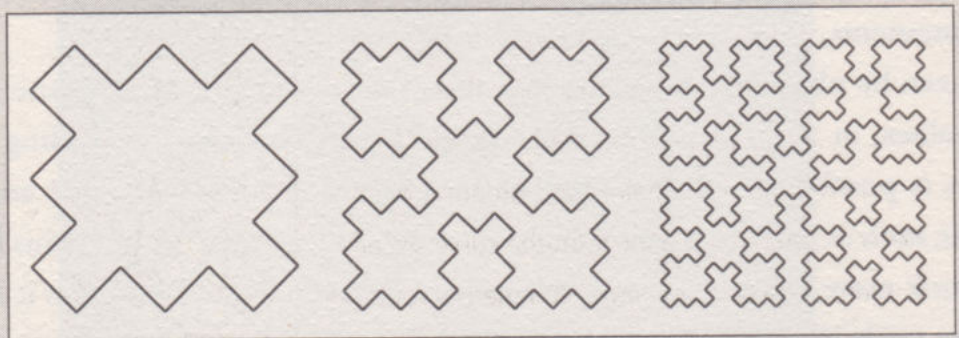
The curve is currently used to provide a highly efficient solution to the problem of the route to be taken by a salesperson, asking what is the shortest path between a series of points on the plane. One possible strategy to solve the problem consists of visiting the points in the same se-

Put another way, the Minkowski dimension is the same as the value of the fraction $\log N(\varepsilon) / \log(1/\varepsilon)$ where ε tends to 0.

The practical idea underlying this formula is applied in the box-counting procedure. The idea is extremely similar to that of the compasses: we have a shape the dimension of which we would like to estimate using a grid of size ε, fixing the value of this latter variable at 1 mm, 1 cm or any other value as a function with of relative size of the shape. We now proceed to count the small squares or boxes that contain part of the shape. Now progressively substitute ε for smaller measurements and count the corresponding boxes. Then, as with the Richardson procedure, make a *log/log* diagram, marking the logarithms of $1/\varepsilon$ on the x axis, and the logarithms of $N(\varepsilon)$ on the y axis. The gradient of the straight line that best approximates the points will give us D_M. It is a systematic procedure that can be applied to any straight line, planar or spacial structure.

The simplicity of the box-counting procedure makes it extremely interesting, however the Minkowski dimension is still missing certain properties that would be desirable from a theoretical point of view. The German mathematician Felix Hausdorff (1868–1942), from the University of Bonn, dedicated himself to studying the theory of measurement and proposed a new definition of dimension which,

quence in which they appear in the Sierpinsky curve. First of all it is necessary to create this curve and draw it in such a way that it covers the points to be visited. If this is not possible, a higher iteration of the curve can be tried. Once it has been possible to draw a Sierpinsky curve that passes through all the points, all that is required is to follow its path, starting from one of the points and ordering them in the sequence in which they appear on the curve. As an example, this idea is used to create the commercial routes for mail or parcel companies. It is also used to minimise the path of a plotter pen when it comes to drawing maps.
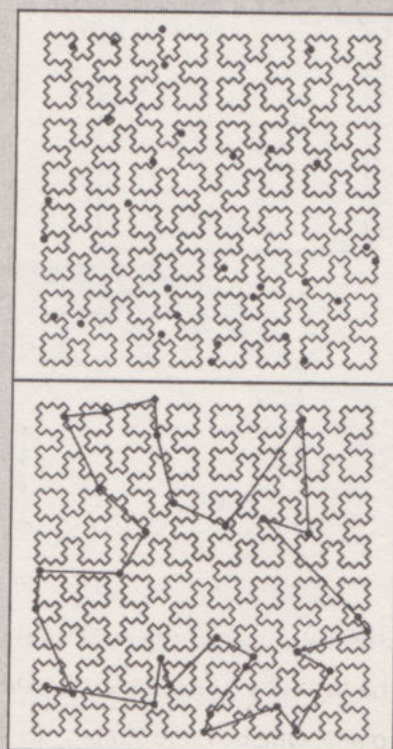
in spite of not being used in practice, is of great theoretical importance and occasionally used for comparing measurements of certain dissimilar sets that nonetheless have the same Minkowski dimension. In general the Hausdorff dimension of a structure is less than or equal to the lower Minkowski dimension.

The dragon curve

Let us now consider a final and surprising curve – the so-called dragon curve. First studied in 1960 by three NASA physicians, Heighway, Banks and Harter, it was later made popular by Martin Gardner in his column on mathematical diversions in the *Scientific American* magazine. The fact that its construction is so simple and its properties so fascinating has inspired many studies in different areas of mathematics. According to Gardner, Heighway constructed it by folding a piece of paper as shown in the figure. We always fold in a valley (or in a 'V'), and the more folds we make, undoing the shape such that all the folds are at a right angle, the closer we get to the dragon curve.
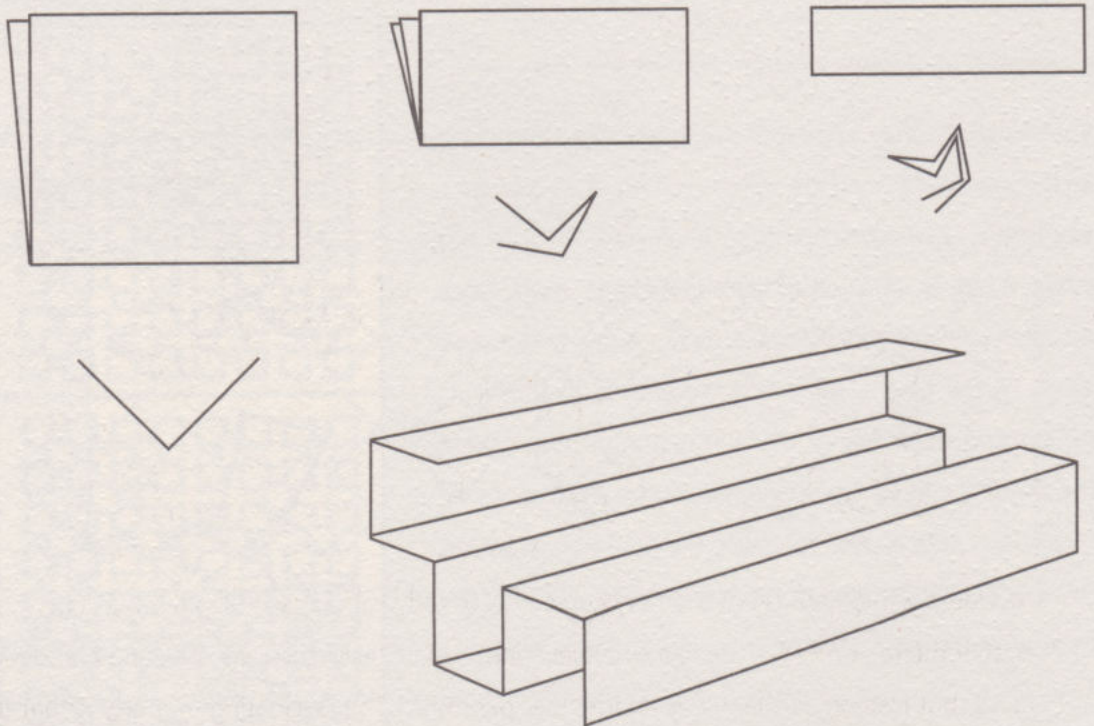
First iterations of the dragon curve. (Source: Maria Isabel Binimelis.)

THE LEVY DRAGON

There are two ways of folding a sheet of paper: into a 'valley' or into a 'mountain'. The dragon curve always folds into a valley. However, combining the two types of folding can give radical changes in the appearance of the curve. There are essentially 16 different ways of creating dragon curves although only five of these are 'essential'. One is known as the Levy curve, also referred to as the Levy dragon or C curve. The first step in its formation is based on a square cut along its diagonal. Using half a square instead of a whole one gives what is known as the Levy carpet.
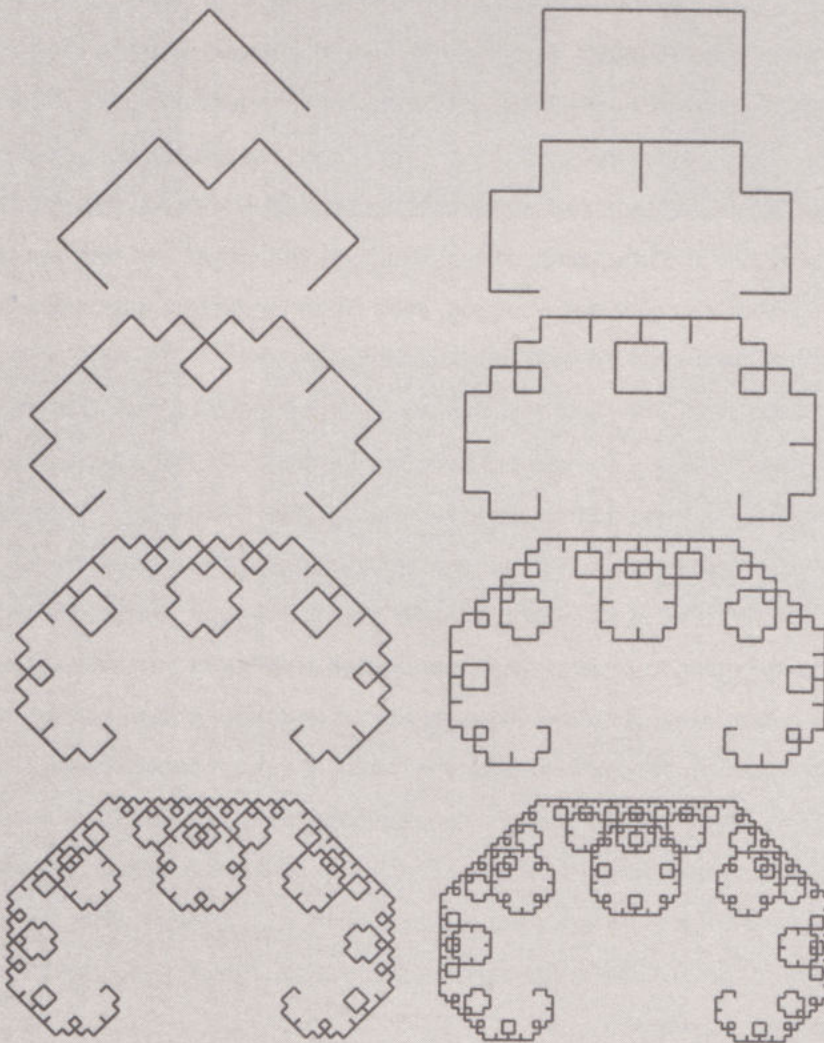
The Hausdorff dimension of the Levy curve is 2, or rather it is space-filling, however we can see that the region of the plane it occupies is not square, as in the case of Peano and Hilbert curves, but instead its shape has an irregular perimeter. The fascinating property is that this perimeter has a dimension of approximately 1.9340.

FRACTALS AND MEDICINE

Physiologists and doctors have only recently begun to quantify the possibilities of the chaotic dynamic in their research, casting doubt on long-standing principles. It was traditionally believed that many illnesses were produced as a result of stress that reduced the "order" of the organism, causing erratic responses or altering the periodic rhythms that are considered to be normal. However, in recent years it has been discovered that the heart and other physiological systems can behave completely erratically when they are young and healthy and, contrary to our expectations, old age and illness are accompanied by increasing regularity.

The human body abounds in fractal-shaped structures: They can be found in networks of nerves and blood vessels, for example. Although the nature of these structures would appear to be diverse, they share common physiological aspects. Fractal branching and folding greatly expand the surface of areas of absorption (such as in the intestine), distribution and collection (blood vessels, biliary tracts and bronchial tubes) and information processing (nerves). Partly as a result of their redundancy and irregularity, fractal structures are robust and resistant to damage. The heart, for example, can continue pumping, even when the system responsible for transmitting cardiac electrical pulses has suffered serious damage.

Listening to heartbeats, the rhythm would appear to be regular. When a person is resting, the intensity of their pulse and the interval between the beats are notably consistent. However, a more detailed analysis reveals that there are considerable fluctuations in the heart rhythms of healthy individuals, even when resting. Represented graphically over the course of the day, the image of this time series looks jagged, irregular and at first sight, completely random. However its self-similarity begins to emerge when represented at different time scales.

The following two series of graphs compare various measurements of cardiac activity for a sick (top row) and healthy (bottom row) person. Those that correspond to the sick person show few oscillations. The graph on the right compares the heart rate at a specific point in time and that corresponding to a certain time delay. In the case of the sick person, the various points are clustered together in a small region. In contrast, healthy people exhibit an erratic aspect with a wide range of values in all the graphs and measurements. Paradoxically, it is the healthy systems that exhibit chaotic behaviour.

The American writer Michael Crichton used the dragon curve in an enigmatic and suggestive way in his novel *Jurassic Park*. Each chapter of the novel is headed by a drawing of the corresponding iteration of the curve followed by a brief comment from Ian Malcolm, one of the characters in the
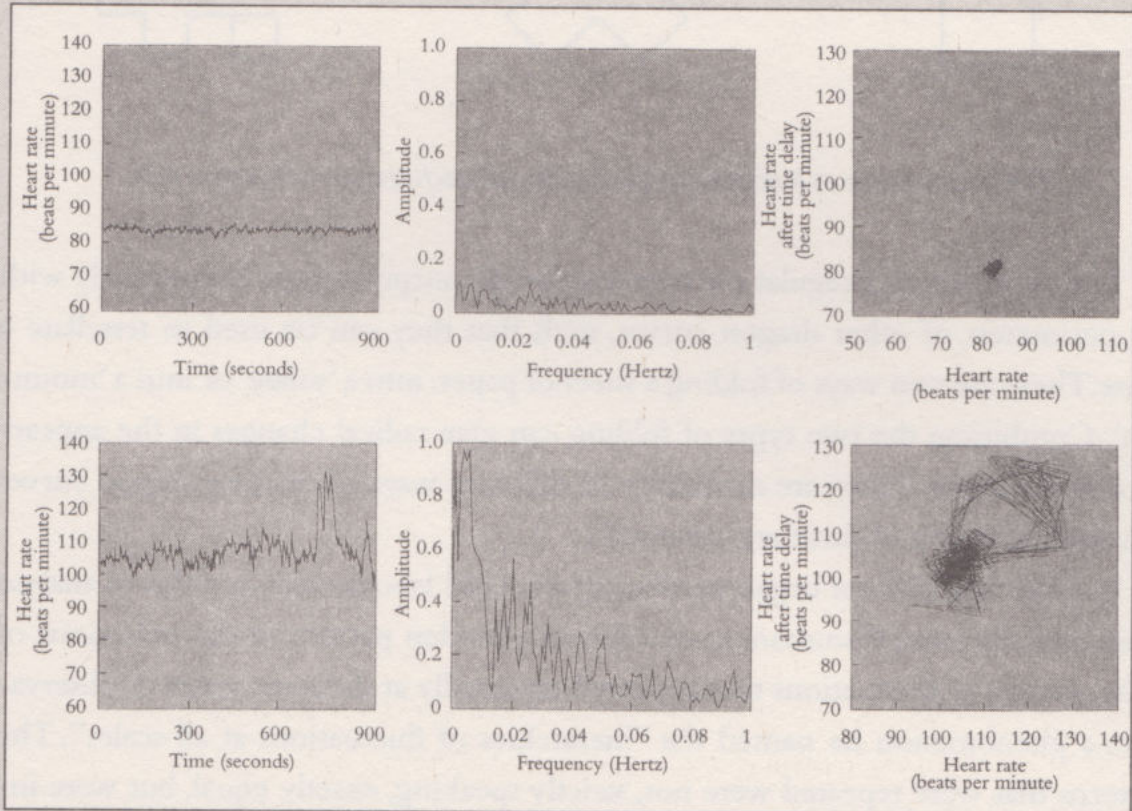
Why do the heart rate and other systems controlled by the nervous system have a chaotic dynamic? Such dynamics can offer a number of functional advantages. Chaotic systems are capable of operating under a wide range of conditions and as such are adaptable and flexible. This flexibility allows systems to adapt to the requirements of a changeable and unpredictable environment.

novel who is a mathematician by profession and a specialist in chaos theory. The story is based on the cloning of dinosaurs using their DNA, and Crichton reproduces the dragon curve as a visual metaphor for the complexity and instability of this process.
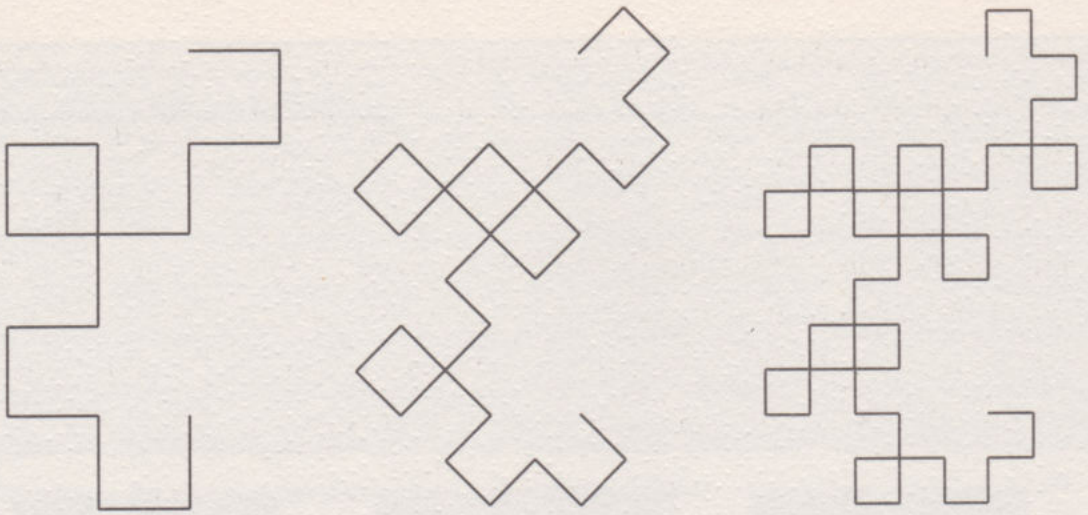
The dragon curve as it appears in the heading of each chapter of Jurassic Park.

The curve has an irregular perimeter although, surprisingly, it fits perfectly with the perimeters of other dragon curves, such that they can be used to tessellate a plane. There are two ways of folding a sheet of paper: into a 'valley' or into a 'mountain'. Combining the two types of folding can give radical changes in the appearance of the curve. There are essentially 16 different ways of creating dragon curves although only five of these are 'essential'.

In 1958, Mandelbrot began working at the IBM laboratories in order to analyse noise and electrical disturbances, and found a hidden pattern in the behaviour of noise, groups of fluctuations that appeared repeatedly at different scales of observation, a phenomenon he named the "hierarchies of fluctuations at all scales". The patterns that were repeated were not, strictly speaking, exactly equal, but were instead statistically similar. However even so, the fluctuations could not be described using known statistical mathematics. Instead of investigating the phenomenon, he began to think of other systems in which it was possible to find similar patterns that behaved in the same way and could not be described exactly using statistical mathematics. Solving these issues led him to suggest methods of observation based on self-similarity and then to the discovery of fractals. Mandelbrot observed that these methods were an extremely powerful tool for the study of phenomena that depend on chance in areas as diverse as geostatistics, economics, physics and medicine.

Throughout this chapter we have seen how, long before Mandelbrot gave them their name, the ideas underpinning the concept of fractals had been present in the minds of certain revolutionary mathematicians. The world as described by Euclidean geometry was confined to cubes, cones and spheres and made up of

straight lines, flat surfaces and circles. Weierstrass, Cantor, Poincaré, Peano, Hilbert, Koch, Sierpinsky and Hausdorff peered beyond and glimpsed fascinating worlds that made use of textures, branches and cracks in the definition of its many and complex objects.

The shapes collected during these journeys defied conventional descriptions and were often branded as mathematical monsters, pathological equations and even a "lamentable plague of functions". However, as a result of their work, the non-conformists have immeasurably enriched mathematics.

Chapter 3

Dalmatians and Dragons: Linear Fractals

Mother Nature did not attend high school geometry courses or read the books of Euclid of Alexandria. Her geometry is jagged, but with a logic of its own and one that is easy to understand.

Nassim Nicholas Taleb, *The Black Swan*

Both in Escher's work and other similar work, we have seen that the concepts of self-similarity and iteration give rise to certain objects which, on account of their approximations of infinity, challenge common sense and lead us to a vertiginous endlessness. Little by little, we will now see how the foundations of a new geometry were laid, based on the concept of self-similarity and the principle of continuity (the latter being introduced by Leibniz). Similarly, in terms of philosophy, an agreement has not yet been reached regarding the scope of the term continuity. In mathematics, the term has changed over time, becoming more specific and being attached to a range of definitions until reaching a precise form. The significance of the concept of continuity in the evolution of mathematics is demonstrated by the simple fact that it has been one of the most studied.

It is often assumed that space and time are continuous, and certain philosophers have maintained that all natural processes occur continuously, whence Leibniz's famous aphorism: *natura non facit saltus*, "nature does not make jumps". Visually speaking, continuous means 'uninterrupted, without separations'. In mathematics, a discipline in which precision prevails, the exact definition of 'continuous' has had a turbulent and tortuous life. For a long time, the definition of a function was also linked to the idea of continuity.[1]

---

1 In 1838 Lobachevsky gave the following definition: "A function of x is a number that is given for each x and that changes gradually together with x. The value of the function could be given either by an analytic expression or by a condition which offers a means for testing all numbers and selecting one from them, or lastly the dependence may exist but remain unknown."

It is difficult to specify how to formulate a statement equivalent to Leibniz's in mathematical terms. Towards the end of the 18th century, it was said that if a function was continuous it meant that infinitesimal changes in the value implied infinitesimal changes in the value of the function. With the abandonment of infinitesimals in the last century[2], this definition was replaced by another, which made use of the more precise concept of a limit.

To say, for example, that a function does not have jumps, lacks precision in mathematical terms. In a first attempt at approximating the correct formulation, it can be said that the graph of this function must be convex (that is to say that it cannot be separated into two open sets that do not intersect). However if we try to be more precise, it is not clear if we should perhaps say that it is arcwise connected[3] instead – that which can be joined by the arcs of a curve.

Nowadays, the concepts of function, continuity and differentiability are precisely defined and constitute an important part of secondary education. A function is now said to be continuous at a given point when its one-sided limits coincide with its value at that point[4]. However, when studying the behaviour of functions, it is often not enough for them to be continuous; other additional properties that make them more pleasing from a mathematical point of view are required as well. One such property is uniform continuity. Generally speaking, this means that small changes in the value of the input give rise to small changes in the value of the function and, moreover, that the size of the changes in the output value depends solely on the size of the changes in the input values and not the value of x (hence the adjective 'uniform'). All uniformly continuous functions are continuous, although the converse is not true. Consider, for example, the function $f(x) = 1/x$ in the domain of real numbers. The function is continuous, but not uniformly so since as x tends to zero, the changes in the value of $f(x)$ grow without limit. In addition to uniformity, there are also many additional properties of continuous functions such as absolute continuity, which is a property of smoothness that is stricter than uniform continuity.

Although the Czech mathematician Bernhard Bolzano (1781–1848) had already anticipated the rigorous formulation of the concept of continuity, his work remained overlooked for many years. Instead, an idea that forms the cornerstone of

---

2 In modern theories such as non-standard analysis, this naming convention is re-used with a new interpretation.

3 Every arcwise connected set is connected although the converse is not true.

4 This definition is equivalent to the so-called 'Cauchy-Weierstrass definition of continuity', or the 'epsilon-delta definition'.
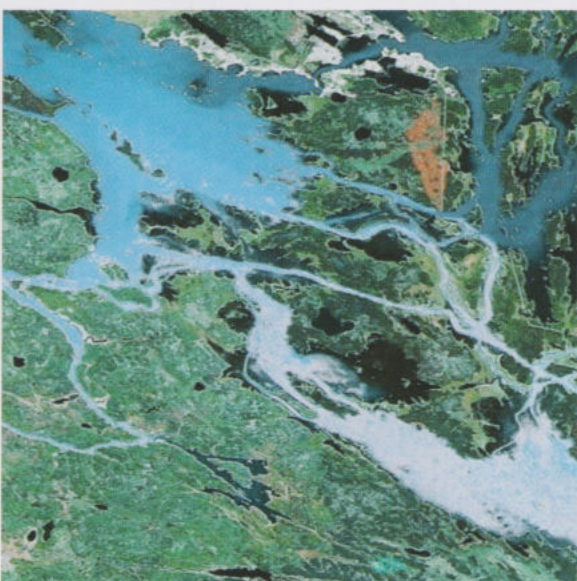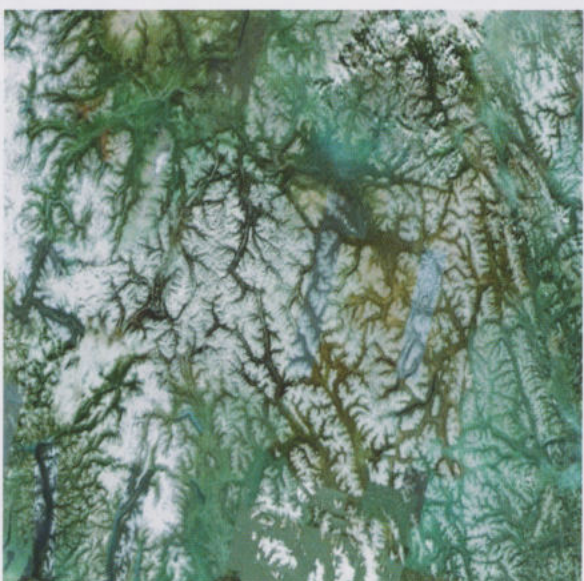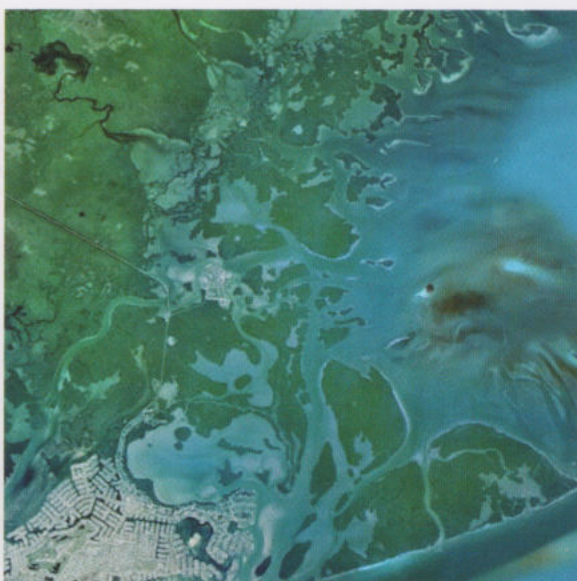
Aerial images taken over areas of the Nile, the Amazon and the Great Lakes. These are surfaces with a highly irregular structure that approximates models from fractal geometry.
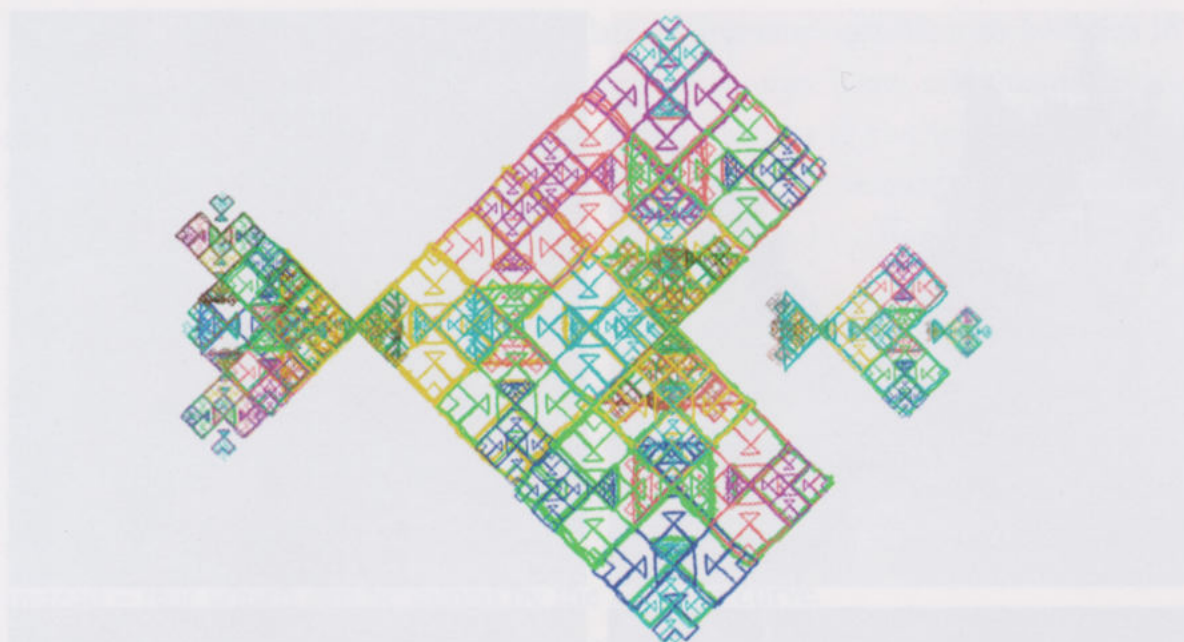
The large fish eating the small fish converted into an attractor of an affine IFS composed of eleven functions. Six are required to cover the body, four for the tail and one more for the smaller fish. Here we see the third iteration represented.
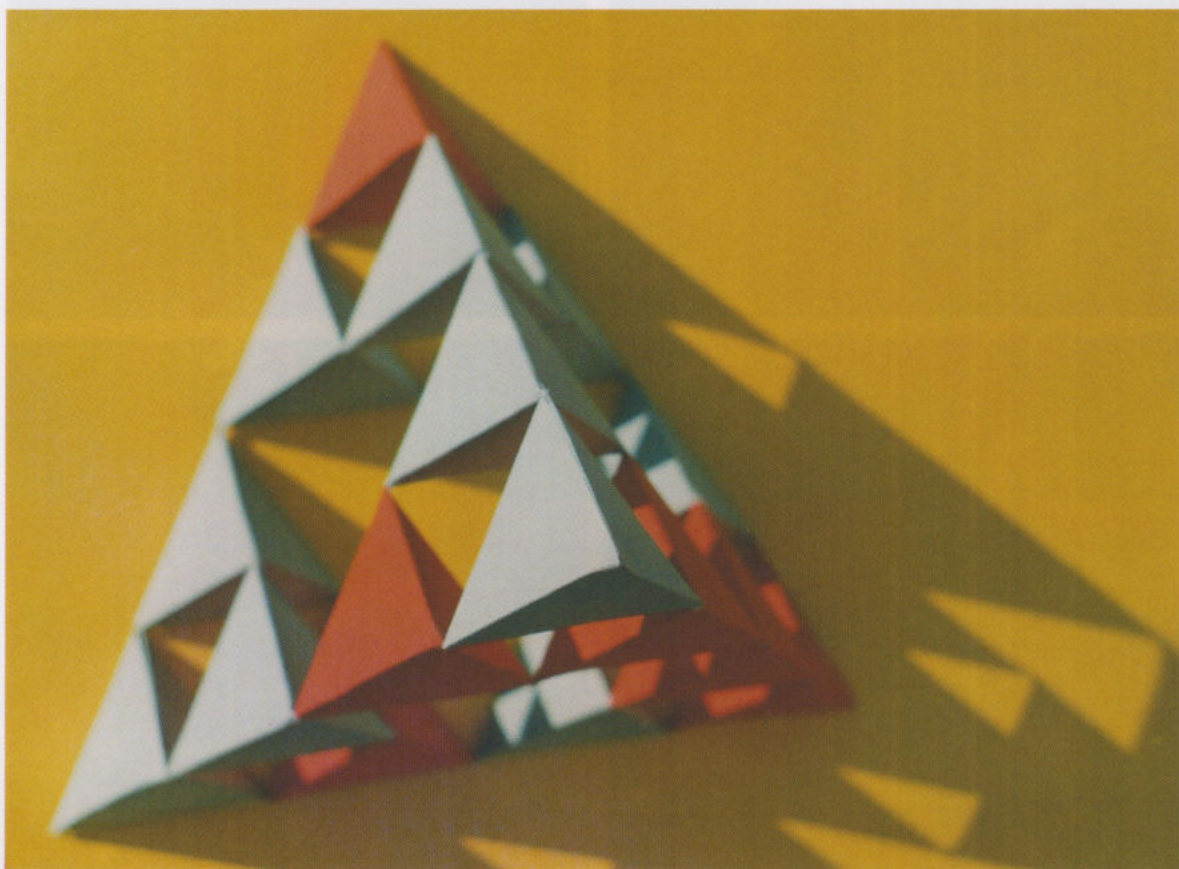
Artistic representation of a sculpture inspired by the Sierpinsky tetrahedron. Its construction is the same as that of the triangle with the same name, only there are now four tetrahedra in 3D space instead of three triangles on the 2D plane.
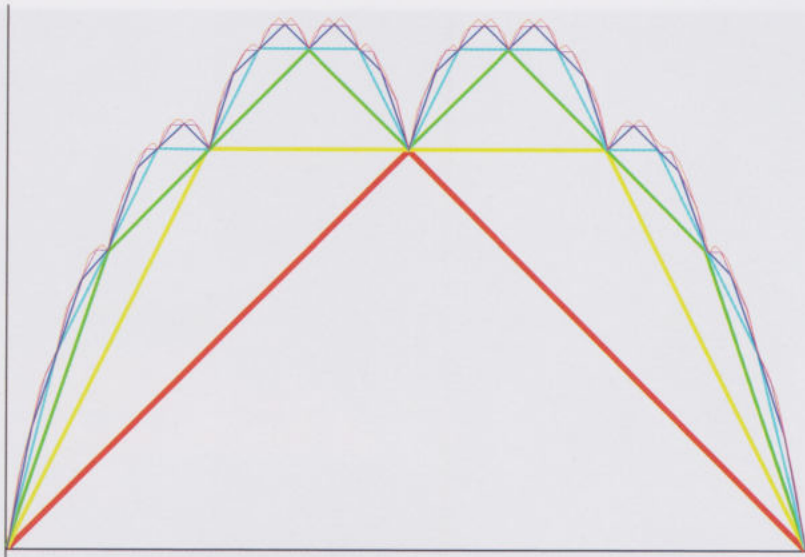
Construction of the Takagi, or blancmange, curve using polygonal lines. Each is constructed based on the previous one by a process known as midpoint displacement, which had been used by Archimedes to calculate the area between an arc of a parabola and its chord.

An algorithm like the one used to create the Takagi function, but in three dimensions and slightly modified with random parameters, has been used to draw these artificial landscapes. The generation of fractal landscapes is used in the production of many feature films.

Certain real phenomena, such as clouds, are better modelled using fractal geometry than Euclidean geometry. The simulation of artificial clouds is achieved using so-called plasma fractals, adapting a dispersion factor that depends on the effect we wish to achieve.

The genetic codes of plants and other living beings are based on the principle of least action, that is to say, the greatest economy is sought when it comes to generating instructions for growth. This explains the abundance of self-similarity in their morphogenesis, and the majority have a fractal structure.
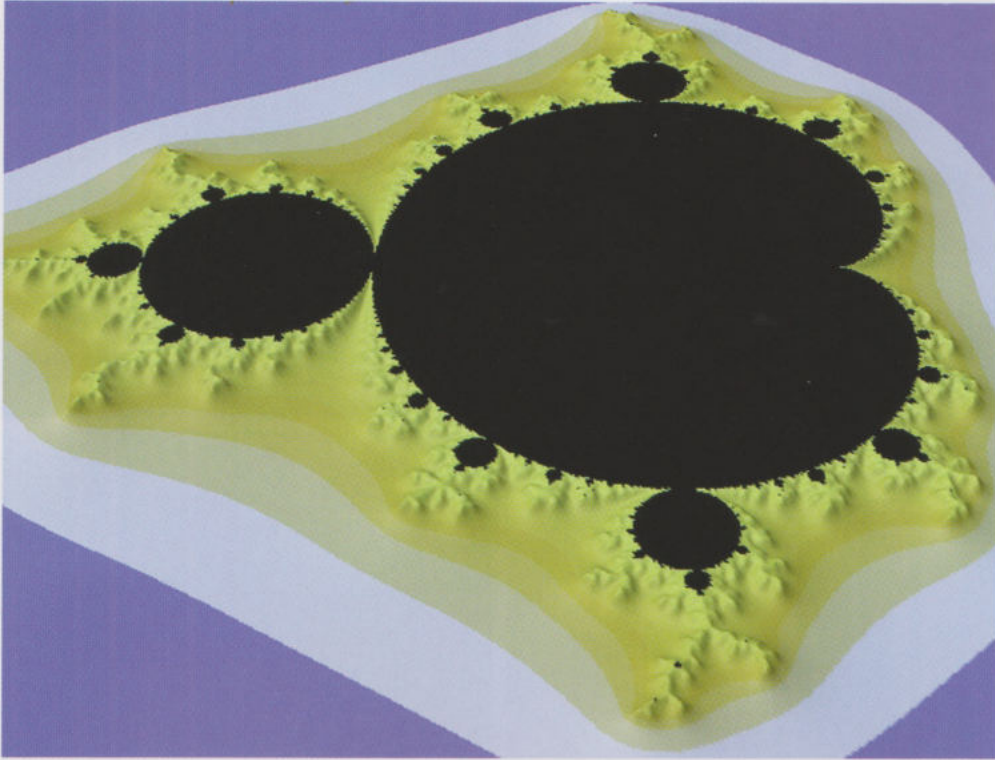
A rendering of the Mandelbrot set in relief using an algorithm based on power curves. The height of a point will depend on the number of iterations required for it to move away from the coordinate centre.
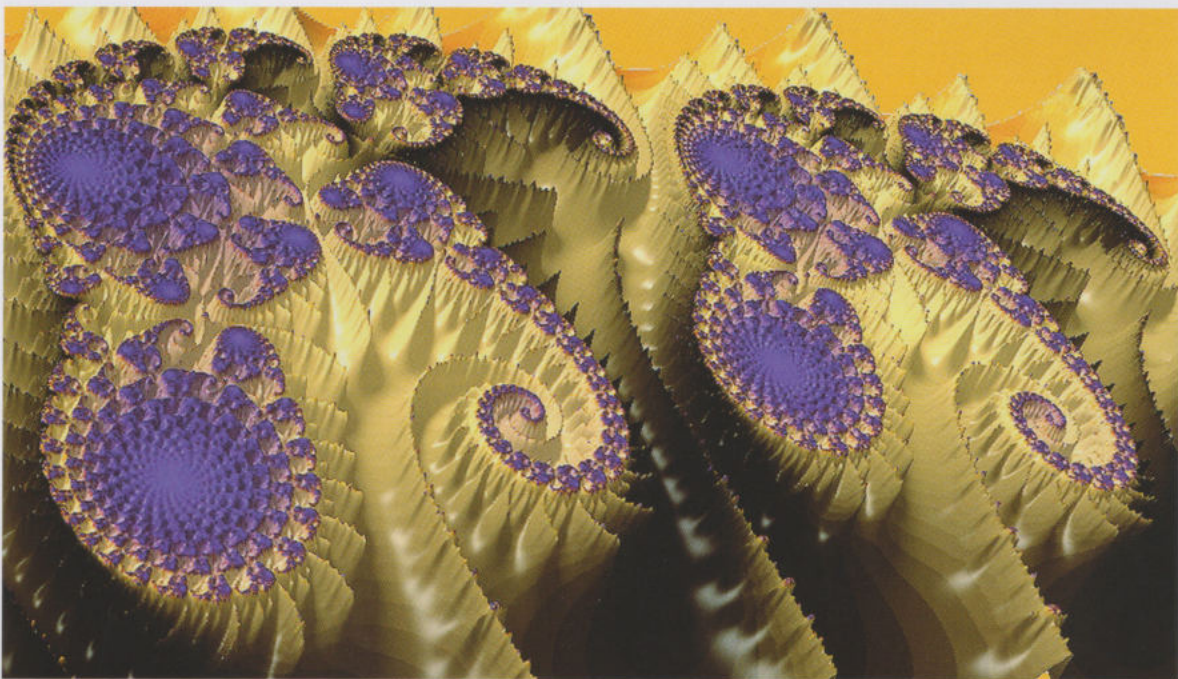
Enlargement in relief of the M set in an area close to the edge of the main cardioid. In relative dimensions, for this enlargement, the whole M set would be as large as the orbit of Jupiter. In this immense universe, we can find fascinating worlds inhabited by elephants, seahorses, snails and an endless number of filigrees.
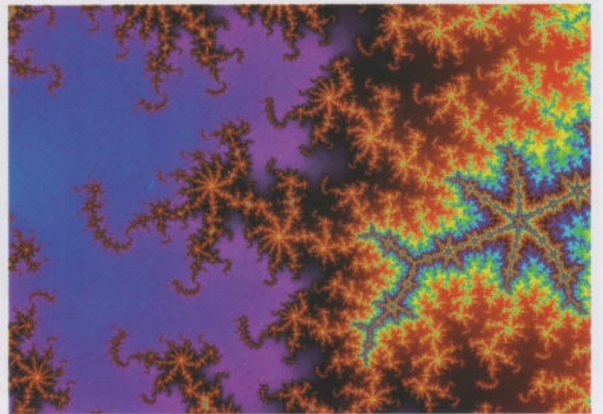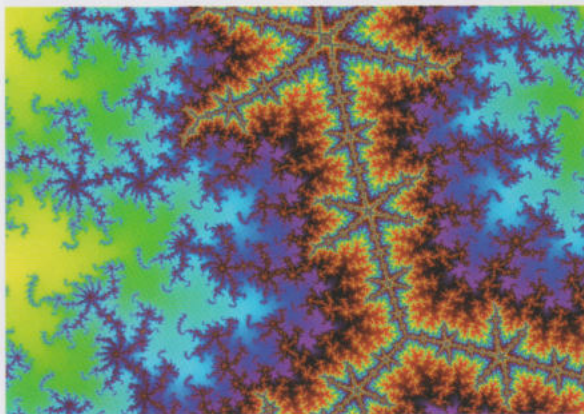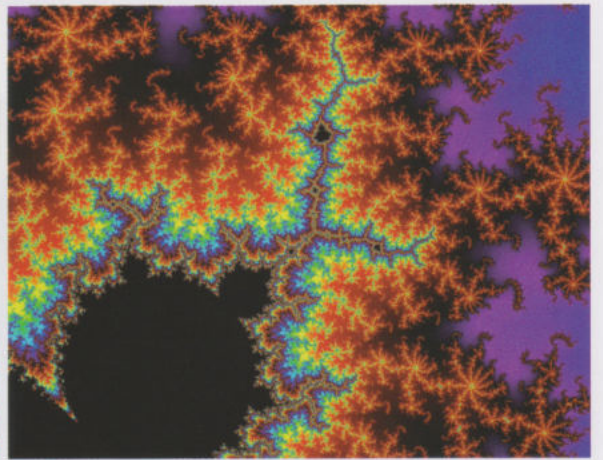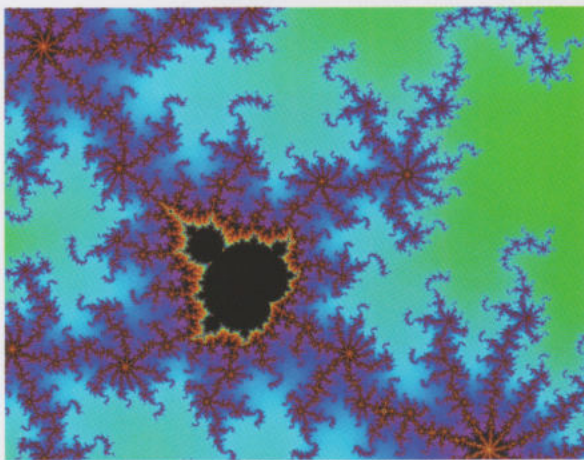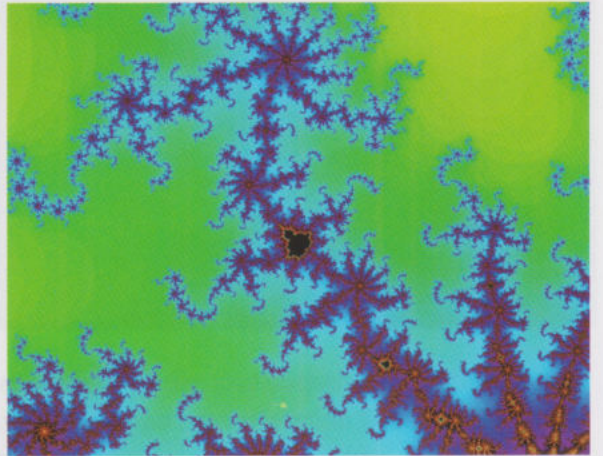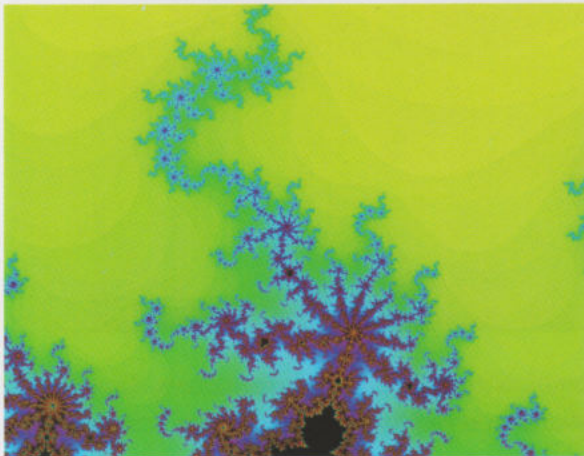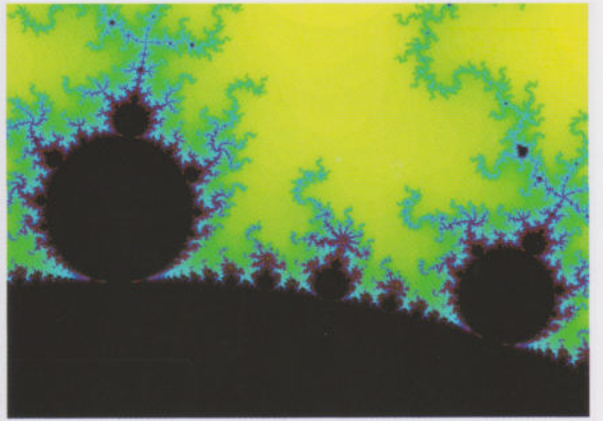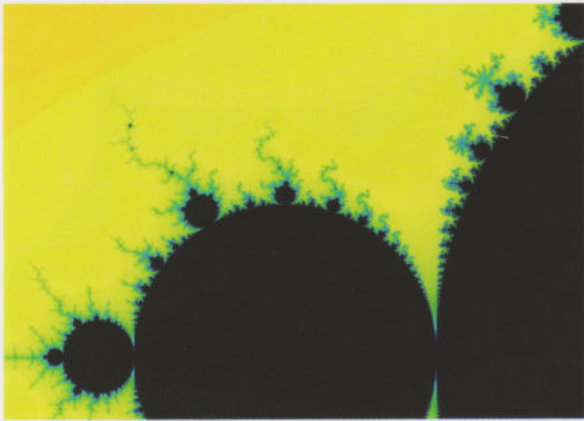
From left to right and top to bottom, successive enlargements of the Mandelbrot set. The centre of each image is approximately the same as that of the previous one.
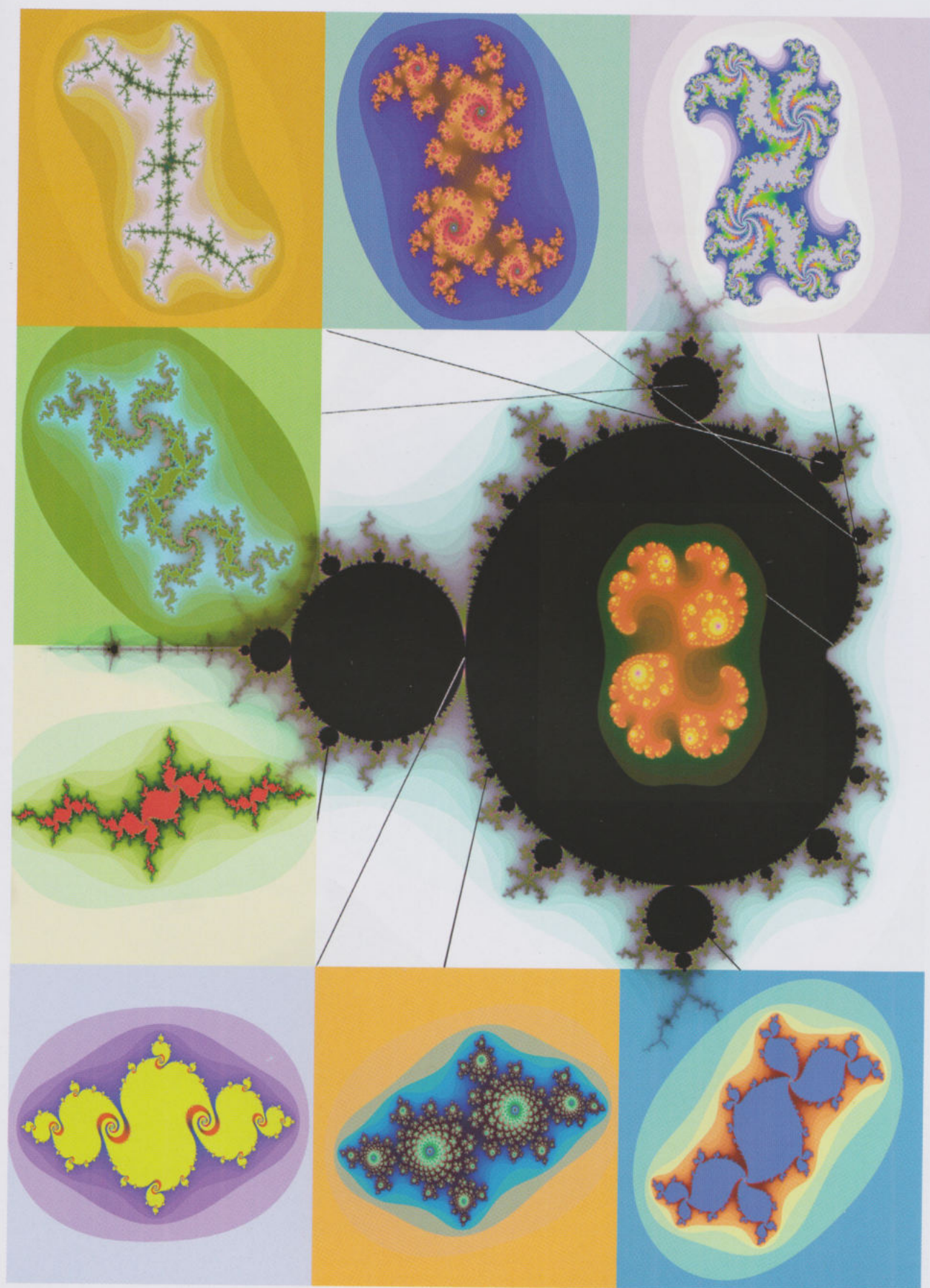
The Mandelbrot set with various Julia sets positioned near the c parameter that generates them.
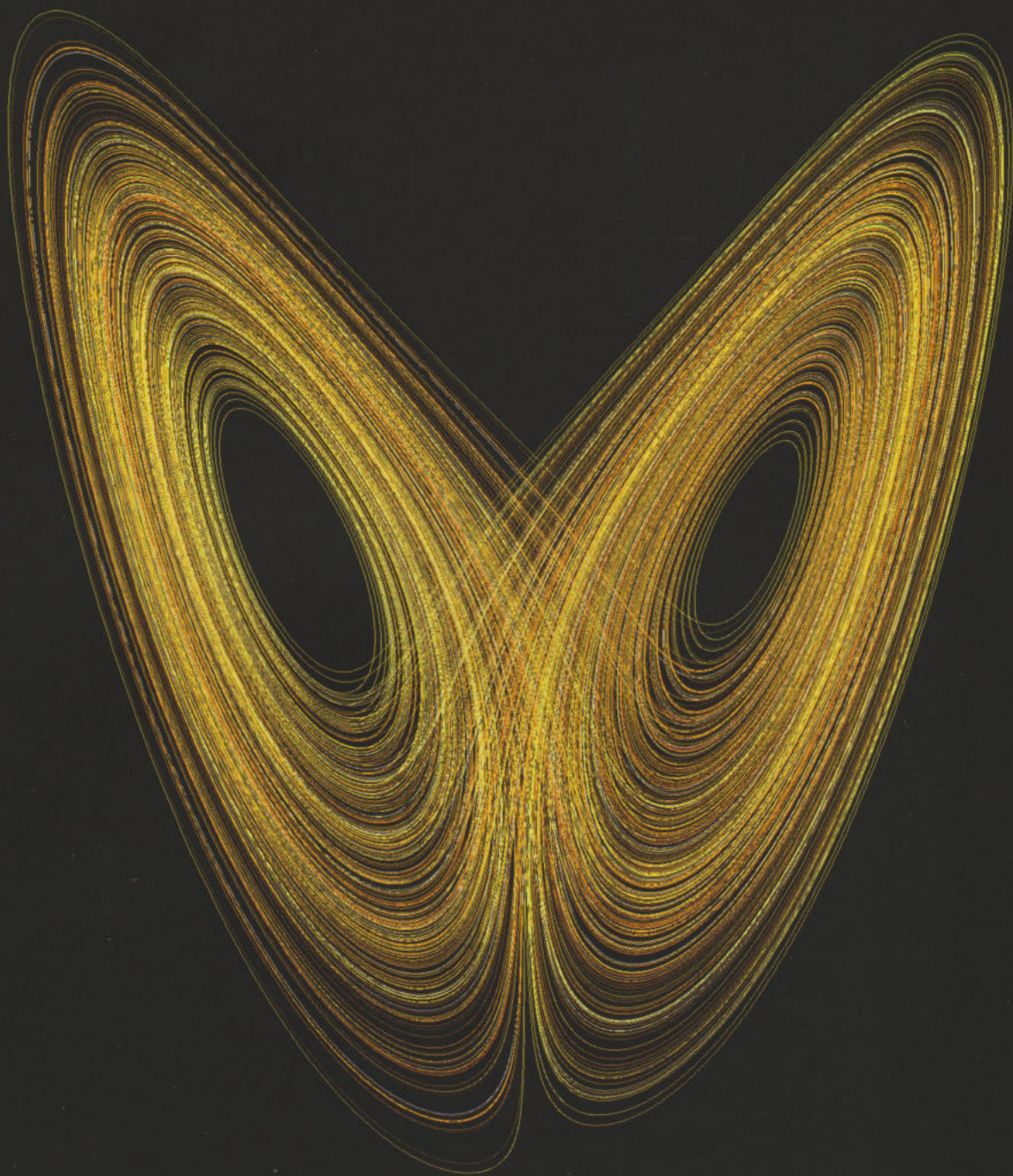
An image of a Lorenz attractor. This represents the orbit of a point in space that evolves according to certain dissipative differential equations which model and simplify the convection of a fluid. The orbit is around the two centres and jumps from one side to another an infinite number of times. In the words of James Gleick, this magical image, similar to the face of an owl or the wings of a butterfly, became the emblem of the first explorers of chaos.

the modern concept of continuity – essentially identical to Bolzano's – was developed by the Frenchman Augustin Louis Cauchy (1789–1857), who described the continuity of a function using the notion of a limit. Hence for Cauchy, a function f is continuous for a if:

$$\lim_{x \to a} f(x) = f(a).$$

The derivative of the function at a given point is the same as the value of the tangent of the curve of the function at that point, and it constitutes an essential part of the calculus, jointly invented by Newton and Leibniz. The idea of differentiability is an intuitive one. Given a certain 'point' on a curve, if it does not have a single tangent, the curve is not differentiable at that point. At the start of the 19th century, the majority of mathematicians believed that a continuous function had a derivative – or rather that the tangent line was well defined – for almost all its points.

In a lecture given at the Berlin Academy in 1872, Karl Weierstrass caused a stir among the mathematics community by showing a continuous function that was not differentiable at any point. The function in question is defined as an infinite sum of sinusoidal functions and depends on two variable parameters, a and b:

$$\sigma(x) = \sum_{n=0}^{\infty} (a^n \cos b^n \pi x).$$

When a is between 0 and 1, the series is continuous. However, Weierstrass proved that it is not differentiable at any point if b is an odd integer and $a \cdot b > 1 + 3\pi/2$. Shortly after, the English mathematician G.H. Hardy proved that it suffices to have $a \cdot b > 1$ and $b > 1$.
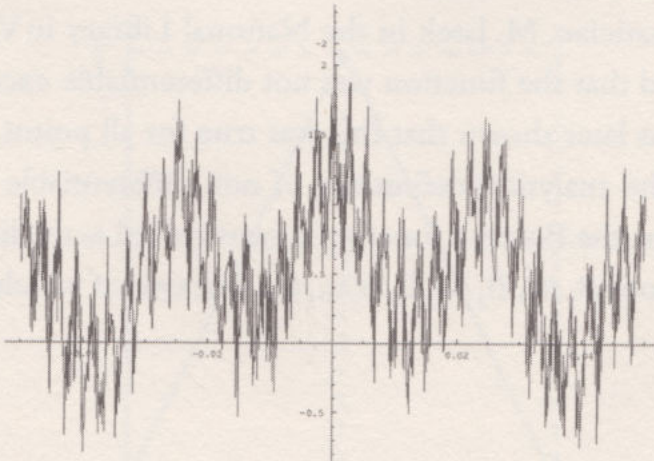

A graphical representation of the Weierstrass function for a = 0.7 and b = 9.

In his writing, Weierstrass mentioned Riemann, who had apparently made use of a similar construction in 1861, although this was not published. Riemann's expression is also a sum of sinusoidals, however this time without parameters, and the index variable n is arranged differently:

$$f(x) = \sum_{n=1}^{\infty} \frac{\sin(n^2 \pi x)}{n^2}.$$

It is not easy to imagine the graph of a function of this nature. The following illustration shows Riemann's function alongside the share price of a given bank over a one year period. This example highlights the extent to which a curve like this can be used to describe real phenomena.
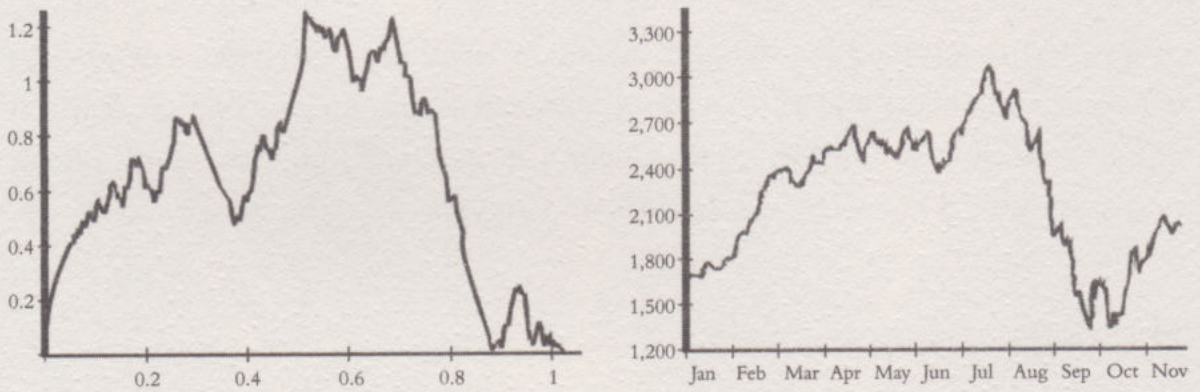
The first known example of a function of this type was not due to Weierstrass's work, but to the aforementioned Bolzano who presented a continuous function that was not differentiable at almost all its points in 1830. The loss of the manuscript delayed its publication until after World War I, when it was discovered by the Czech mathematician, M. Jasek in the National Library in Vienna. In the paper, Bolzano proved that the function was not differentiable except for a set with measure zero. It was later shown that this was true for all points.

In contrast to the analytical definitions of non differentiable functions, which we have already seen, the Bolzano function is constructed as the limit of a sequence of polygonal functions $B_1(x)$, $B_2(x)$, $B_3(x)$..., the first two of which are shown in the following graphs.
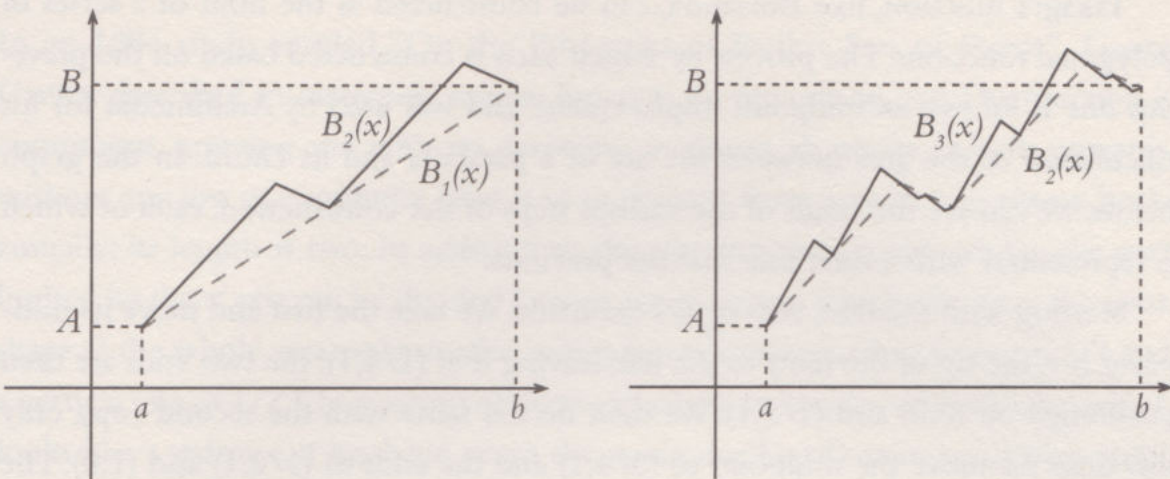
Source: Maria Isabel Binimelis.

From Weierstrass onwards, examples of continuous, non-differentiable functions began to proliferate. A phenomenon that was initially thought to be uncommon was shown to be normal, to such an extent that we now know that continuous functions that are also differentiable are comparatively scarce.

Puddings and tents

In 1903, Teiji Takagi (1875–1960) provided a more straightforward example than Weierstrass of a continuous function that was nowhere differentiable. Its analytical definition is extremely similar. In terms of its graphical representation, the cosine is essentially substituted for a 'tent' function in an inverted V shape as can be seen in the following graph:
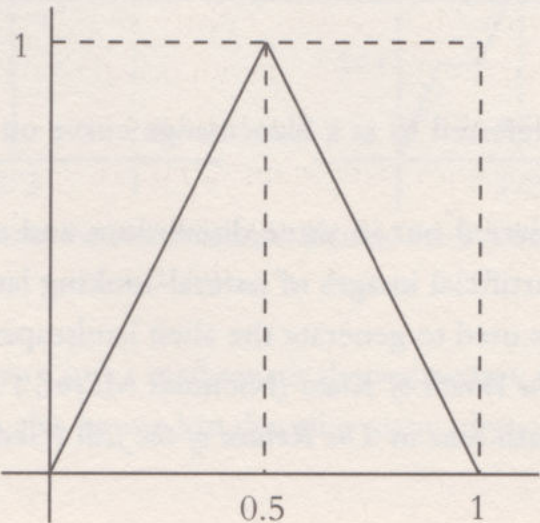
Takagi's function, like Bolzano's, can be constructed as the limit of a series of polygonal functions. The process by which each is constructed based on the previous one is known as 'midpoint displacement' and was used by Archimedes for his calculation of the area between the arc of a parabola and its chord. In the graph below, we can see the result of the various steps of the construction, each of which is represented with a finer line that the previous.

Starting with the tent, two copies are made. We take the first and move its midpoint (i.e. the tip of the tent) to the left, leaving it at $(1/4,1)$; the two ends are then positioned on $(0,0)$ and $(1/2,1)$. We then do the same with the second copy, only this time we move the midpoint to $(3/4,1)$ and the ends to $(1/2,1)$ and $(1,0)$. The resulting polygonal line constitutes the second step of the construction. To draw the third step, we proceed in a similar way, repeating the process until reaching the limit, the curve that runs over the top of the drawing.
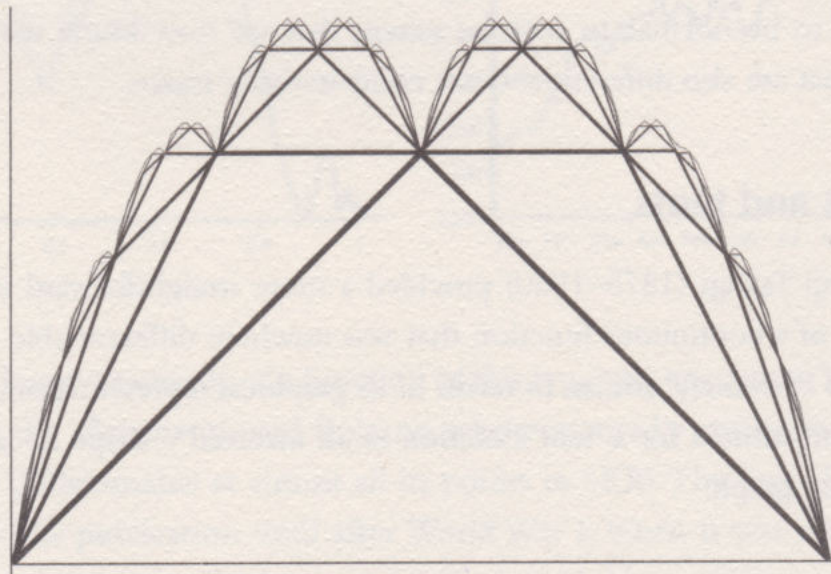

This curve is also referred to as a blancmange curve on account of its resemblance to the dessert.

If this process is carried out in three dimensions and random displacements are used, it results in artificial images of natural-looking landscapes. As an example, this procedure was used to generate the alien landscapes in large blockbusters such as *Star Trek II, The Wrath of Khan* (Nicholas Meyer, 1982), and the exterior of the incomplete Death Star in *The Return of the Jedi* (George Lucas, 1983).

The Devil's staircase

In an 1884 study entitled "On the Properties of Perfect Sets of Points", Georg Cantor described an extremely strange function on the unit square. The function is continuous, growing and with no derivative at almost all points. It leads upwards, without any loss of continuity (that is to say jumps) from zero to one, always horizontally; its length is two. In addition to this, the function is self-similar: the part limited by the x axis can be divided into six equal copies, which also have the same shape as the whole area, only smaller, with a horizontal reduction factor of 1/3 and a vertical one of 1/2. On account of these enigmatic properties and its shape, which looks like a staircase, it has been given the name the 'Devil's staircase'. Other staircases have also received this name, but this is the most apt and appears to be the first to have been described in this way.

The Devil's staircase is obtained by the following iteration. Taking the diagonal of a square, on the central third of the x axis, we raise a segment to a height of 1/2, as shown in the drawing. In the second step, we proceed in a similar manner, lowering the diagonal to the height of 1/2. The diagram below shows the first two iterations. If we apply the algorithm an infinite number of times, we obtain the staircase.
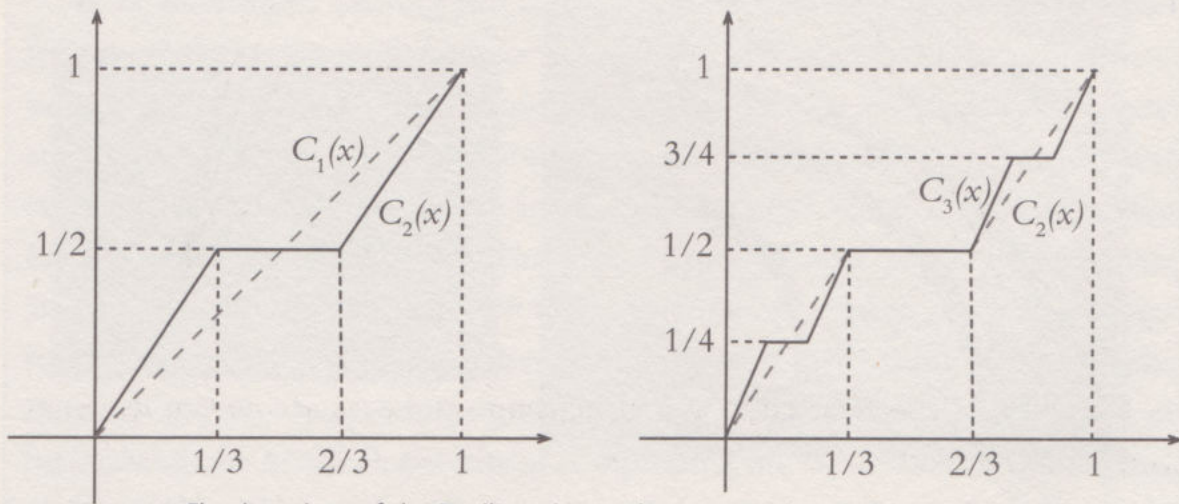

First iterations of the Devil's staircase. (Source: Maria Isabel Binimelis.)

The Devil's staircase is not just a mathematical construction with interesting properties, but its shape can also be used to describe many physical systems.

DIABOLICAL LENSES

Contrary to what happens with conventional or refractive lenses, diffractive lenses focus light as a result of diffraction, a phenomenon that occurs when light interacts with the physical structure of the lens in the form of concentric rings with a different thickness and/or transmittance. There are new diffractive lenses known as 'Devil's lenses' which offer the user benefits such as increasing the depth of focus and reducing chromatic aberration. In spite of their satanic name, the creators of this discovery confirm that there is nothing evil or occult about this discovery, not even close: "The reason behind such an eye-catching name is connected to the profile of these lenses which have been designed according to a fractal structure known in mathematics as the Devil's staircase."

The lenses have the property of being intrinsically multifocal (with many focus points between them) such that the distribution of intensity presents a fractal structure. The multifocality of a lens allows the focus points corresponding to different wavelengths of an image to superimpose themselves to a greater extent, thus generating a sharper final image. This same property of multifocality makes it possible to increase the depth of field, or rather extend the region of space in which an image of acceptable sharpness is obtained.
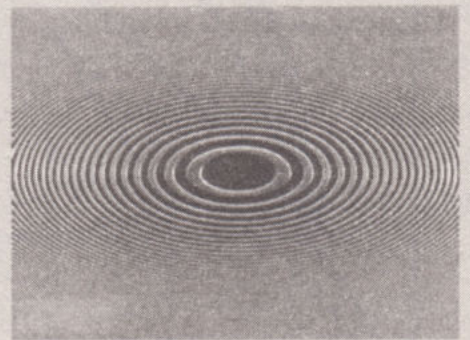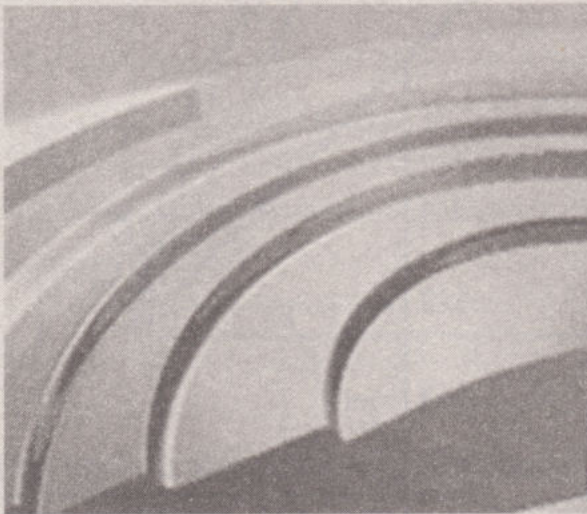
The photographs show two diffractive lenses viewed under a microscope in such a way that it is possible to see their structure, made up of many different concentric rings. Their design is based on a fractal structure. Multifocal lenses of this type are used to correct presbyopia, and their intraocular version to replace the crystalline lens in cataract surgery.

What do the following have in common: dust, snowflakes and sponges?

The simplest transformations that can be carried out on an object on a plane are so-called 'similarity' transformations. As their name indicates, the object is transformed into another similar one, that is to say, its shape does not change, although its position, size and orientation may alter. Translations, dilations, reductions, rotations and reflections are all similarity transformations.

We say, for example that Koch's curve, which was discussed in the previous chapter, is self-similar because it is made up of a number of sections that are similar to the whole, specifically four. The first section or segment to the left of the curve is obtained by simply reducing the whole curve by one third and positioning the left end to the left of the whole.

The second section is obtained by again reducing the whole curve by a third, however this time it is rotated 60° with respect to the horizontal and its left end is positioned to the right of the first segment. During this process, we have made use of translations, reductions and rotations – which are all similarity transformations.

We can proceed in the same way with the Sierpinsky triangle. This time however, we do not even have to apply a rotation, all that is required is to translate and reduce by one third. The same goes for the Cantor set, the Sierpinsky carpet, the Menger sponge and even the dragon curve.

If, in addition to rotating and making symmetries, we add two new transformations, one that allows us to deform the width and height of a shape in different proportions, and another that allows us to rotate the coordinate axes by different angles, we obtain the set of planar affine transformations. The first transformation converts a square into a rectangle and the second, often referred to as shear, transforms a square into a rhombus. The fractal structures that can be obtained using affine transformations are referred to as self-affine. Among the fractals belonging to this type is the well-known Barnsley fern, attributed to the Briton Michael Barnsley and which reproduces itself repeatedly. It is made up of four affine transformations, one of which consists of compressing the width to 0 and thus obtaining the stalk, whereas the other is a shear transformation (crop of the image) which gives the 'branches'.

These transformations can be used to construct a range of different fractals, which are given the common name linear fractals or Iterated Function Systems (IFS). The systems are obtained by means of a series of transformations applied to a set. According to the formulation presented in Barnsley's popular book *Fractals Everywhere*, IFS

The Barnsley fern and its various affine transformation.

make up system of iterative functions. A shape known as the 'attractor' arises as the result of the iteration of the transformations. This is the shape the fractal tends to when the system of transformations is iterated a sufficient number of times. Although it may seem surprising, the attractor does not depend on the initial pattern that is chosen. All of the fractals that have been previously shown can be constructed based on this set of transformations.

We are going to make use of the IFS to describe the dragon curve from the previous chapter. In spite of its apparent complexity, the construction of this curve requires just two transformations. To show that it does not depend on the initial set, we shall construct it based on two different initial patterns, a segment and a shape.

In the case of the segment, and in order to facilitate our explanation, we shall start with the unit segment. It is first reduced by a factor of $1/\sqrt{2}$ and rotated by 45° in an anticlockwise direction. Its left end is then positioned on the point (0,0). It is then reduced by a factor $1/\sqrt{2}$ once more and rotated 135°, again in an anti-clockwise direction, positioning its left end at (1,1).

It can be noted that in the first step the transformations of the unit segment are continuously linked. This is due to the fact that we have calculated the reduction factor in such a way that it is formed by the ends of the diagonal of a square. If we apply these transformations to the resulting shape of the first iteration, we obtain the second, and so on. See how the transformations quickly begin to look like the dragon curve.
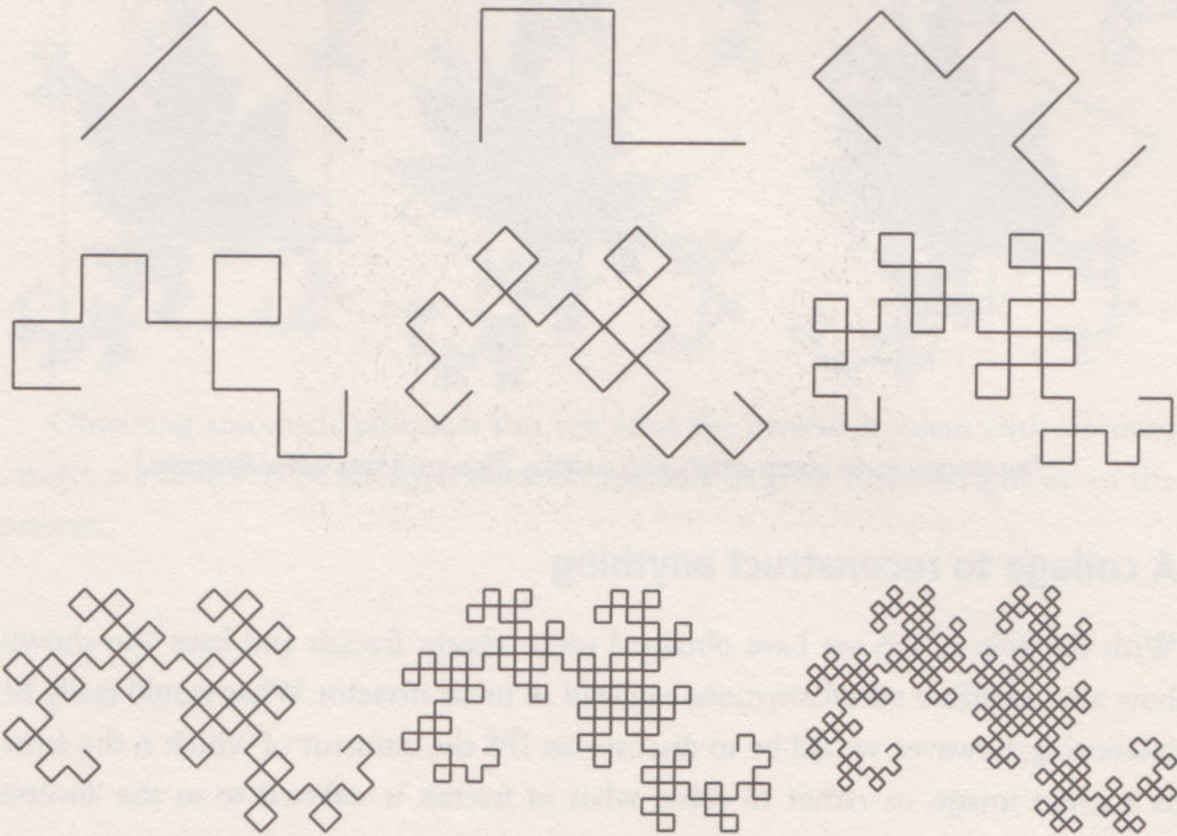
The dragon curve constructed using a segment. (Source: Maria Isabel Binimelis.)

Now to show a construction based on a shape, we shall use a dalmatian dog to produce 101 dalmatians... or maybe even more. The construction of the dragon curve based on the initial dalmatian is identical to the one used to obtain the line version above.
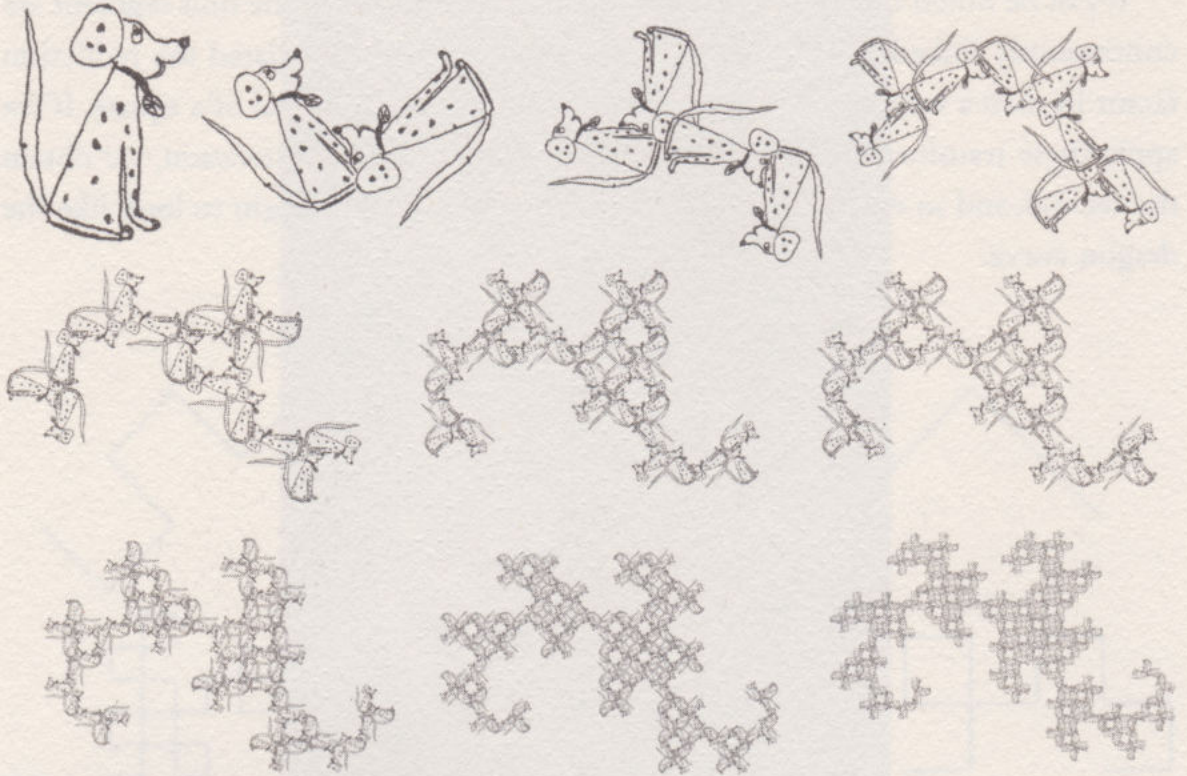
*The dragon curve constructed using a shape. (Source: Maria Isabel Binimelis.)*

A collage to reconstruct anything

With the help of IFS we have obtained some 'classic' fractals and have also shown how a set of affine transformations can lead us to an attractor. What would really be interesting, however, would be to discover an IFS the attractor of which is the same as a given image, or rather to solve, what in fractals is referred to as the 'inverse problem'.

In this respect, one of the most important advances has been the collage theorem devised by Barnsley in 1985. Let's assume that we have a set L and an associated IFS. In order to know to what extent the IFS approximates L, we calculate the image of L for each of the functions and join them together in a single figure. The difference between the image we obtain and the original figure gives us an idea of how that IFS approximates L.

For example, assume that we have a maple leaf and we wish to approximate it by means of an IFS. Given that the collage theorem does not tell us which IFS to choose, we have to base the procedure on our intuition and try to make a *collage* with various copies of the original leaf. To facilitate this process, in our example we will only look for affine functions and attempt to solve the problem using three of

them. The following images show the deformation of each copy with respect to a rectangle with a different shape. Once overlapped, we position them as if it were a transparency on the original leaf. The shape in black represents the parts that have been covered; on the other hand those that are in shades of grey represent part of the error which, according to the collage theorem, determines how far the final attractor of the IFS is from the maple leaf.
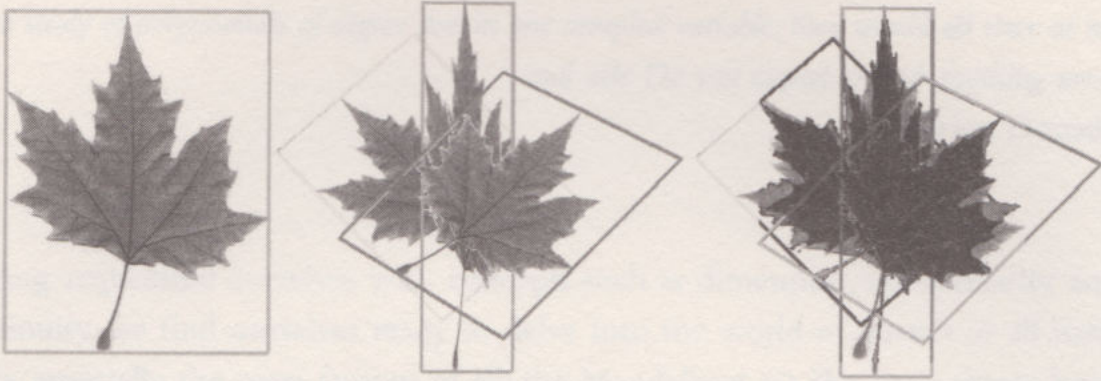
Obtaining automatic processes that can solve the inverse problem with arbitrary images continues to be an open research problem of great interest right up to the present.

Chapter 4

Order Disguised

In 1980, whenever I told my friends that I was just starting with J.H. Hubbard a study of polynomials of degree two in one complex variable, they would all stare at me and ask: Do you expect to find anything new?

Adrien Douady

Having acquainted ourselves with concepts such as dimension, self-similarity and continuity, we find ourselves ready to delve into the world of fractals in all their glory, especially the most famous of all: the Mandelbrot set. We must always keep in mind the fundamental lesson learnt throughout this book: even simple rules for forming fractals can generate extremely complex structures. This principle holds not only for the shapes we have seen until now or those that we will study in this chapter, but also for a large quantity of natural phenomena. Fractal geometry offers analogies and models that will perhaps allow us to discover a universal law for the cosmos in the future. Should such a law exist, an old enemy must lie at its heart: chaos.

Did Mandelbrot discover the Mandelbrot set?

The Mandelbrot set, also known as the M set, has a range of remarkable properties, but perhaps the most mysterious is that its infinitely complex structure is based on extremely simple principles. These can be understood by anyone who knows how to add and multiply, even though studying the set may require billions of sums and calculations to be carried out. This is why the M set was not discovered until the advent of modern computing.

As we shall see further on, the basic theories that made the discovery of the M set possible were developed during the second decade of the last century by the French mathematicians Gaston Julia (1893–1978) and Pierre Fatou (1878–1929). In 1918, Julia published various articles related to complex numbers that

studied the properties of certain sets which, at that time, could not be visualised and would later go on to be referred to as Julia sets.

In 1978, the Frenchman Adrien Douady (1935–2006) and the American John Hubbard (1945–) managed to use purpose-built software to obtain the first blurry, poor quality images of Julia sets. A year later, Mandelbrot published his own images that were developed at the IBM laboratories. The first image of the M set is dated to 1981 and is a result of joint work by Robert Brooks and Peter Matelski.

COMPLEX NUMBERS ARE NOT SO COMPLEX

Real numbers are a way of labelling each of the points located on the real straight line in a unique and ordered manner since each number corresponds to a point and each point has its label. There are rules for adding, subtracting, multiplying and dividing them. So-called complex numbers operate in a similar manner, since they also allow us to label points, although these are not to be found on a line but on a plane referred to as the complex plane.

There are three common ways of labelling a complex number: Cartesian, arithmetic and polar. The number is characterised using the variable z. In its Cartesian form, a complex number z is made up of two coordinates: one real and the other imaginary. The imaginary coordinate is the vertical one and its unit is i. This i is referred to as imaginary since it is the square root of -1. As a result of the peculiarity of this imaginary number, the numbers that contain it became known as complex. Although they are still referred to by this name for historical reasons, they are in fact not so complex, as we shall see. Indeed, thanks to them it has been possible to simplify many statements and theories. In order to visualise them on the plane, a coordinate system is defined whereby the horizontal axis is used to signal the value of the real component and the vertical axis for the imaginary. We can write a complex number in Cartesian, arithmetic and polar form:

$$z = \left(\frac{1}{2}, \frac{\sqrt{3}}{2} \right) = \frac{1}{2} + \frac{\sqrt{3}}{2i} = 1_{60°}.$$

Thus, z is a complex number in which the real part has a value of 2 and the imaginary part a value of $\sqrt{3}$, the same as saying that it has length 1 and forms an angle of 60° with the horizontal. To add two complex numbers together, all we need to do is add the real and imaginary parts separately. For example, if $z_1 = (-2,4)$ and $z_2 = (3,1)$, the sum is $(1,5)$. The diagram shows that the sum is nothing more than the point which falls on the diagonal of the parallelogram made up by the vectors indicated by the numbers. For multiplication, the following rule is

The rigorous mathematical study of the M set also began with Douady and Hubbard, who proved that the set was connected and compact and that its interior consisted of an innumerable set of components. Finally, they were responsible for writing the canonical version of its formula in the form of a complex quadratic equation $z^2 + c$.

In his time Mandelbrot claimed that the main problem he had to overcome when studying the set that now bears his name was programming the algorithm to

followed: i^2 is substituted for -1, since, as we have seen, i is the root of -1:

$$z_1 \cdot z_2 = (-2 + 4i) \cdot (3 + i) = -2 \cdot (3 + i) + 4i \cdot (3 + i) = (-6 - 2i) + (12i - 4) = -10 + 10i.$$

The last point is the one which, geometrically speaking, corresponds to the vector the angle of which is the sum of the two angles and the length of which (referred to as the modulus) is the product of the lengths of the two initial numbers.
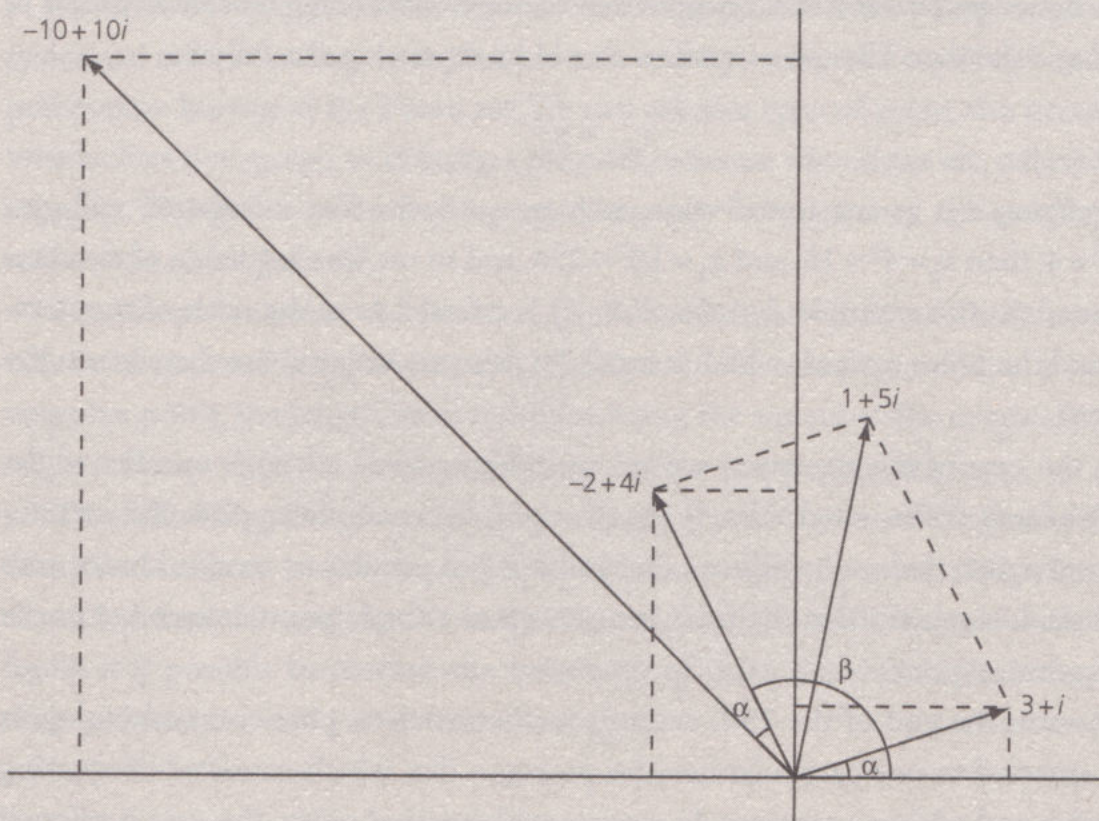
allow him to see it. In his book *Fractals: Form, Chance and Dimension* he recognises his debt to the pioneering work by Julia and Fatou and states:"I proceeded in a way abhorrent to theoreticians, I studied and dissected images that are unforgettable, using the computer like a microscope, a primitive tool of 1980."

Impulsive-compulsive calculations

The most important conclusion to be drawn from all this work is that a simple formula can give rise to complex results. This discovery marks something of considerable importance for science as a whole, as we shall see further on.

In an iterative procedure, the result of a calculation is re-used as an input. The idea is based on taking a number on which an operation is carried out, repeating it with the result and continuing to do so indefinitely with the results that are obtained each time. Formally it is said that an iteration is carried out and is represented in general terms as:

$$x_{n+1} = f(x_n).$$

To better understand this, imagine that the operation being repeated consists of squaring a number. Then the iteration would be represented as follows:

$$x_{n+1} = x_n^2.$$

Applying this to any initial value, such as, $x_0 = 2$, the first calculation will give $x_1 = 2^2 = 4$; then $x_2 = 4^2 = 16$, and $x_3 = 16^2 = 256$, and so on. The sequence of numbers generated (in this example: 2, 4, 16, 256,…) is referred to as the orbit of the iteration, and the point towards which it tends (in this case infinity) is referred to as the attractor.

In the case of the operation we are considering here, raising a number to the power of two, if the initial value is less than one, for example $x_0 = 0.5$, the attractor is zero; if $x_0 = 1$, the result is always one and it is not possible to escape. Under such circumstances, we say that the orbit is made up of a single point, referred to as the fixed point.

Towards the end of the 19th century, mathematicians, physicists and biologists were attracted to one particular iterative process – that which consisted of squaring a number and adding a constant. In precise mathematical terms, this case is referred to as the real quadratic family. The scientific community's interest in this family

stemmed from its relationship to a series of diverse theories which gradually came to be known as chaos theory.

Prisoner points or escaping from the labyrinth

Julia and Fatou were pioneers in the study of iterations with complex numbers and their results were the basis on which subsequent fractal geometry was constructed. Among other aspects, Julia and Fatou studied the behaviour of complex numbers where the iteration consisted of squaring them and adding a constant. This is expressed symbolically:

$$z_{n+1} = z_n^2 + c,$$

where z is a complex number and c, a constant that is also complex. Simply put, this iteration states: take a number, multiply it by itself, add the constant c and repeat this over and over again with each result. The sequences of orbits that are generated are sequences of complex numbers that only depend on the values of each starting point and the constant c.

In 1906, Fatou applied this iteration to all the points of the complex plane and generated orbits that largely tended towards infinity, except for a well defined set of points now known as the Fatou set. We can call this type of point, the iteration of which does not escape to infinity, a prisoner, whereas the others are referred to as escapees. The points of the border separating one from another, the guardians, are referred to as the Julia set.

Let us now examine the iteration of the constant $c = (0,0)$ in a little more detail. In its geometric representation, the square of a complex number is another complex point, the length of its modulus being the square of the original and the angle it forms with the horizontal being double the original angle.

The following table gives the values of z, z^2, z^4, z^8… up to z^{32} for three different complex numbers. Inside the unit circle (i.e. a complex number with a modulus that is less than the unit), on the unit circle and, finally, outside the unit circle. In the figure it is possible to observe the behaviour of these three examples in terms of their geometry.

		inside the circle		on the circle		outside the circle	
		modulus	angle	modulus	angle	modulus	angle
z_0	z	0.8	10°	1.0	10°	1.5	50°
z_1	z^2	0.64	20°	1.0	20°	2.25	100°
z_2	z^4	0.4096	40°	1.0	40°	5.06	200°
z_3	z^8	0.1678	80°	1.0	80°	25.63	40°
z_4	z^{16}	0.0281	160°	1.0	160°	656.90	80°
z_5	z^{32}	0.0008	320°	1.0	320°	431,439.89	160°
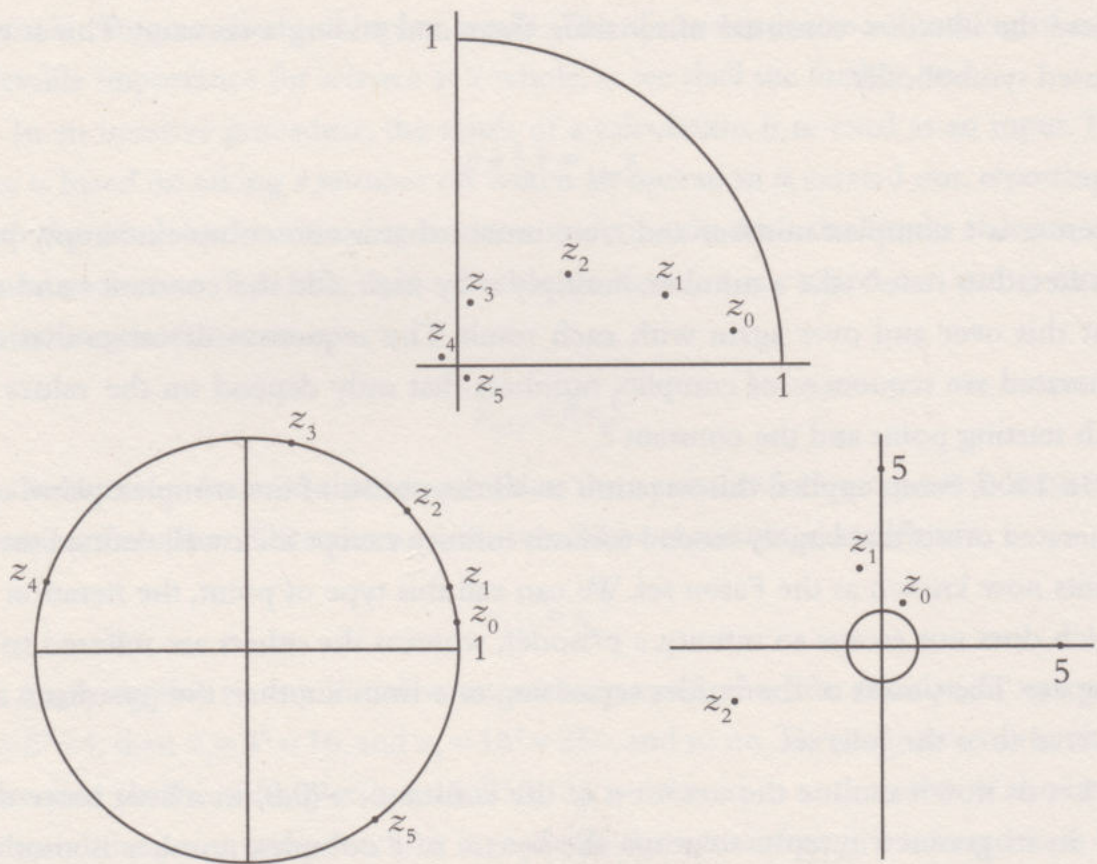
The table above gives the calculations for three types of orbits. The one on the left tends towards the origin of the coordinates; the one in the middle is maintained on a unit circle; and the one on the right tends to infinity. The three figures below the tables show representations of these three orbits on the complex plane.

We can see that the point inside the circle converges towards the origin, the one outside escapes to infinity and the one that is on the circumference always remains on it. The larger the modulus of the initial number, the faster it escapes. This divides the complex plane into two regions: one with the points of the unit circle, which are prisoners, and those outside it, the escapees. In this case, the Julia set is the unit

circumference that corresponds to the guardian points. We can also note something that subsequently becomes extremely important. The Julia set is invariant under quadratic iteration, or rather any orbit that begins inside it continues inside it.

Note that there are two fixed points under the iteration that are (0,0) and (1,0). In this case (0,0) is a fixed attractor point since all the points inside the circle will tend towards it. It is also said that the inside of the circle is the basin of attraction for the point (0,0). However (1,0) is a fixed point of repulsion since there are points close to it, such as (1.01, 0), that escape to infinity.

If we consider infinity as another point on the plane, designating it with the symbol ∞, we can say that the point ∞ is fixed and that its basin of attraction includes all the points outside the unit circle.

This is the most simple example of the Julia set, with a unit circumference. However, it shares the typical properties of the majority of Julia sets: it is the border of the basin of attraction of (0,0) and ∞, and the dynamic between the points is chaotic.

The special case $z_{n+1} = z_n^2$, often written as $z \to z^2$, is the portal to a fascinating and beautiful zoo of Julia set fractals.

The use of a computer is required to visualise other Julia sets, such as $c = 0.5 + 0.5i$. In theory, for each point on the plane, we would have to check if its orbit tends to infinity or zero. In practice, this is impossible and in order to display the points of the Julia set, we must make use of alternative algorithms.

The following shape includes a table with the orbit of certain points and a drawing of the Julia set, corresponding to $c = 0.5 + 0.5i$.
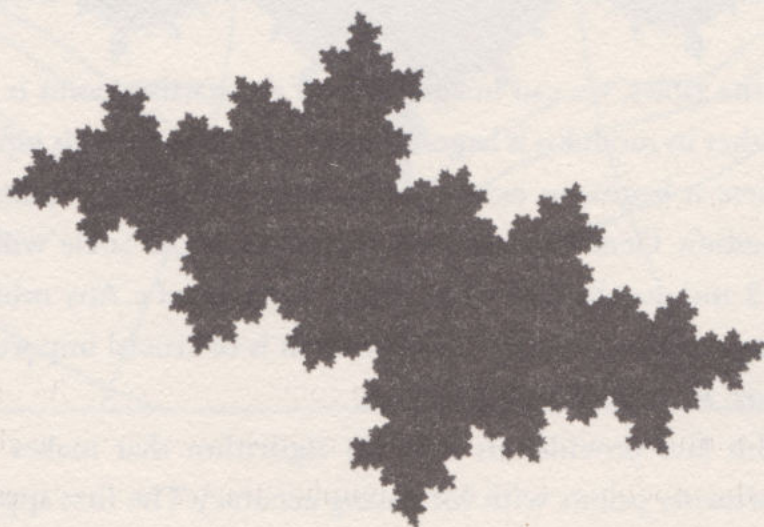
	Orbit 1		Orbit 2		Orbit 3	
	x	y	x	y	x	y
z_0	1.00	0.00	−0.96	0.03	−1.27	0.50
z_1	0.50	0.50	0.43	0.44	0.87	−0.77
z_2	−0.50	1.00	−0.51	0.88	−0.34	−0.85
z_3	−1.25	−0.50	−1.01	−0.39	−1.12	1.07
z_4	0.82	1.75	0.37	1.30	−0.41	−1.90
z_5	−2.90	3.33	−2.04	1.46	−3.93	2.04
z_6	−3.26	−18.92	1.52	−5.47	10.79	−15.52
z_7	−347.46	123.68	−28.01	−16.27	−124.77	−334.48

Three orbits which escape.

	Orbit 1		Orbit 2		Orbit 3	
	x	y	x	y	x	y
z_0	0.000	0.000	0.500	−0.250	−0.250	0.500
z_1	−0.500	0.500	−0.313	0.250	−0.688	0.250
z_2	−0.500	0.000	−0.465	0.344	−0.090	0.156
z_4	−0.688	0.250	−0.371	0.355	−0.456	0.013
z_{100}	−0.473	0.291	−0.393	0.290	−0.438	0.218
z_{200}	−0.394	0.278	−0.411	0.271	−0.409	0.290
z_{400}	−0.408	0.276	−0.409	0.275	−0.409	0.276
z_{500}	−0.409	0.275	−0.409	0.275	−0.409	0.275

Orbits of certain points of the quadratic iteration with $c = 0.5 + 0.5i$. In the top table all the points escape. In the bottom one, all tend to a certain fixed attractor point (−0.409, 0.275).

Looking at the tables, we can imagine that if the starting point is far away from the centre, or rather its modulus is large, the orbit will tend towards infinity. But what is the value where it begins to escape? Fortunately there is an optimal way of answering this question. Generally speaking, the radius of the circle will be the maximum between 2 and the absolute value of the modulus of c. Any orbit that exceeds this circle will be certain to escape. This last result is of crucial importance in determining the M set, as we shall see further on.

Verifying this fact provides us with an algorithm that makes it possible to round up the prisoner points with increasing accuracy. The first approximation to the prison with $c = 0.5 + 0.5i$ will be a circle with radius 2. Programming this algorithm in such a way that it operates on the pixels of the screen, at a rate of

one point per pixel, we have a large set of points (depending on the precision we wish to obtain), which is nonetheless finite. The computer calculates the first iteration for each of the points and renders each of those that escapes the circle of radius 2 on the first iteration in a specific colour while leaving the remaining points black. The border of this black set will be the second approximation to the prison. Based on the black points (their number will decrease with each step), we can calculate the second iteration, painting the points that escape from the radius 2 in another colour and leaving the remainder in black.

This procedure can be repeated for each of the points in black (the number increasing with each iteration) until the change in the black set is no longer visible on the screen. This algorithm, often referred to as the escape-time algorithm, gives the following Julia set for $c = -1$:
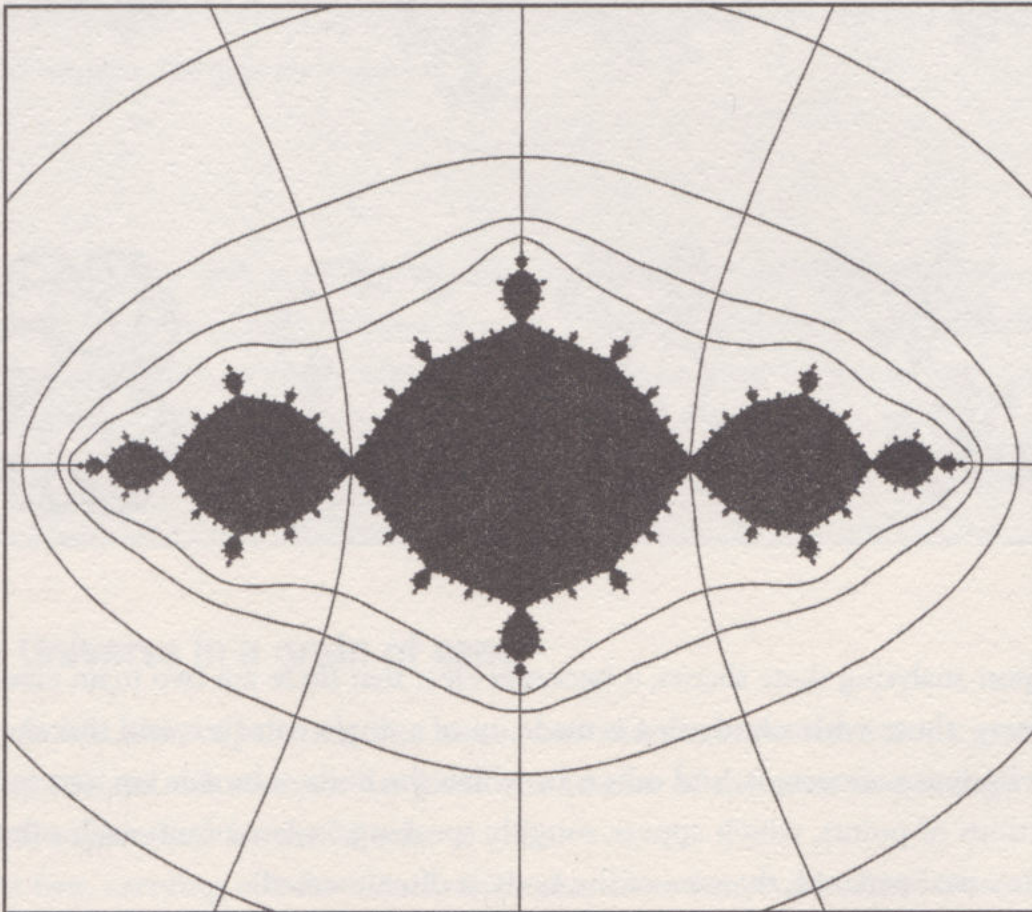
The Julia set corresponding to $c = -1$ with successive approximations (the lines that surround it) calculated using the escape time algorithm.

Different values of c produce different Julia sets:
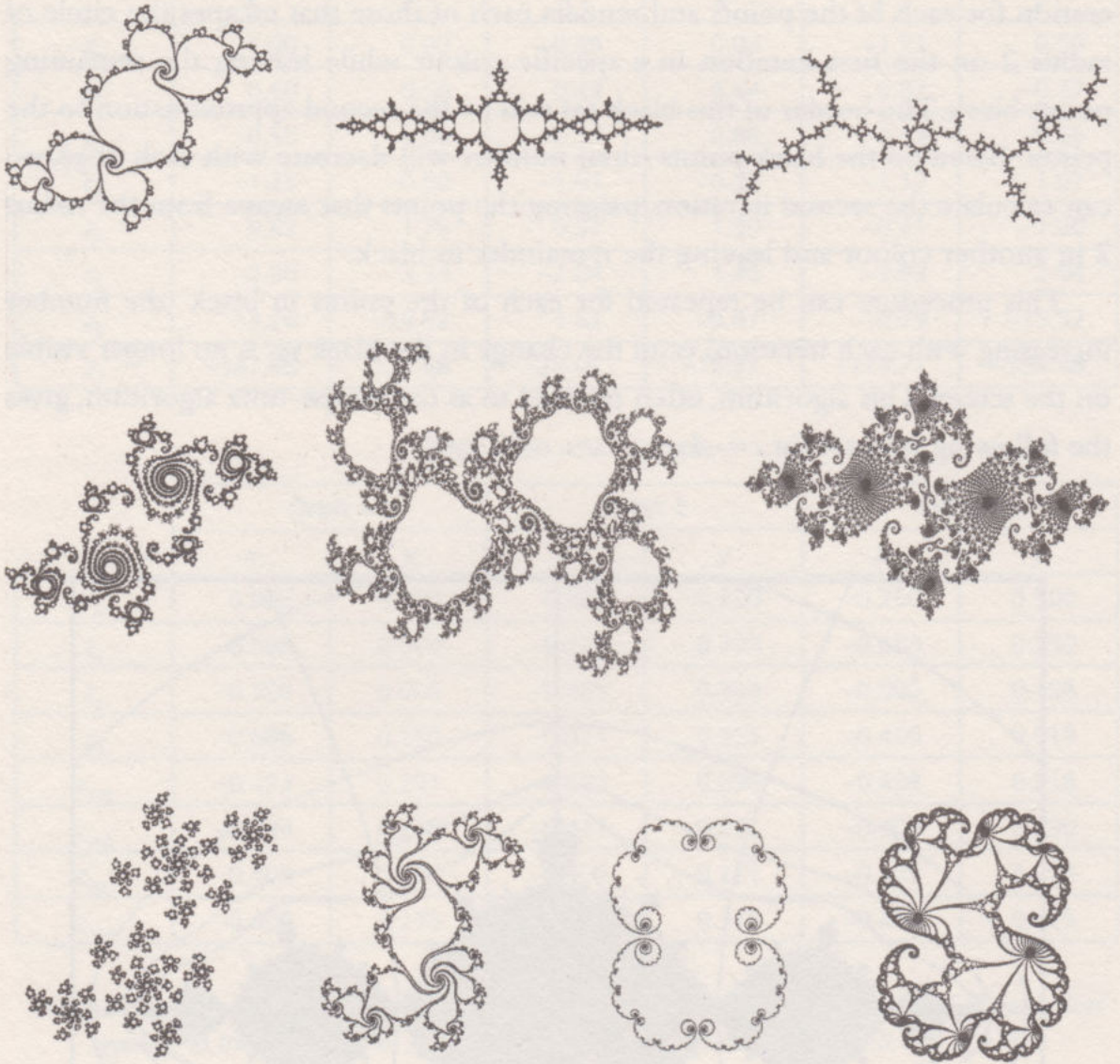
Upon analysing these shapes, it becomes clear that there are two main classes of Julia sets – those with a body that is made up of a single part (it is said that the area of the body is connected), and others in which the body is broken up into infinite collections of points, which appear, roughly speaking, isolated from each other (in mathematical parlance, the area of the body is disconnected).

This geometric distinction serves as the basis for the possibility of separating the values of the constant c, which we shall refer to as the complex parameter from now on, into two well distinguished sets: those that give rise to connected figures in the iteration $z_{n+1} = z_n^2 + c$ and those that give rise to disconnected ones.

118

THE CHAOS GAME

The method for the calculation of successive approximations to the prison using the escape-time algorithm is extremely slow. In order to obtain a sufficiently detailed image of the Julia set, a different algorithm is often used, one which has received the suggestive name of 'the chaos game'. In the previous chapter, we spoke of certain transformations referred to as affine which, when applied iteratively, resulted in a linear fractal. For the task at hand, however, we require transformations that give a Julia set when applied iteratively. However these transformations cannot be affine because Julia sets are not linearly self-similar. Instead, let us begin with the idea that when the points close to the Julia set (and outside it) are iterated with the transformation $z \rightarrow z^2 + c$, they tend to infinity. That is to say, it is as if the Julia set were a repulsor. If the inverse operation is now considered, that is to say the transformation that takes the previous point as iterated, the Julia set now changes from a repulsor to an attractor. What inverse transformations do we need to apply? Let w be the next point of the quadratic iteration, hence $w = z^2 + c$. If we want to obtain a previous iteration, we must separate z from the equation. This gives two solutions:

$$z = +\sqrt{(w - c)};$$
$$z = -\sqrt{(w - c)}.$$

The chaos game works in the following way. An initial arbitrary point is taken and the two images calculated according to previous results. The process with all the obtained points is repeated and they are shown on the screen. Each time we repeat the process, we will get nearer to the true appearance of the Julia set.

The Universe in a grain of sand

The classification of Julia sets into connected and disconnected is not a gratuitous dichotomy, and this research has given rise to one of the most fascinating mathematical objects known: the M set.

At first, carrying out the classification may seem intractable, since we could be forgiven for thinking it necessary to analyse all the points of all the Julia sets for each variation of the parameter c, of which there is an infinite number. However Mandelbrot took advantage of a theorem that was proved independently by both Julia and Fatou in about 1919, according to which the orbit of the 0 plays an

essential role in deciding whether a Julia set is connected. Specifically, the theorem states that if the orbit escapes to infinity, the Julia set appears fragmented; otherwise, it is connected. The implications of this theorem are enormous, since it is now only necessary to apply the iteration to a single point, $z_0 = (0,0)$, and thus determine the nature of the Julia set that is obtained applying the iteration to the complex plane.

The above result provides an accurate and convenient way to determine whether or not a Julia set is connected. However, how do we know when the orbit of (0,0) diverges to infinity? The answer is already known: the orbits diverge if at some point they exceed the circle of radius 2 and that of radius $|c|$.

Mandelbrot used this property of quadratic iteration and plotted the values of the constant c that give rise to connected Julia sets. He then found that this collection of the values of c has a surprising structure when represented on the complex plane.

Roughly speaking, the Mandelbrot set can be described as a cardioid (a heart shape) with an infinite number of tangential discs, from which, on account of its size, one on the top stands out. Enlarging this particular disc, it becomes clear how the structure is joined to other 'similar' structures by means of filaments, and although it would seem that there are points scattered here and there, the full set is a single piece, or rather is connected.

The inside of the set has a dimension of two. However, while the topological dimension of its border is one, in 1991, the Japanese mathematician Mitsuhiro Shishikura proved both to his own surprise and that of others, that its Hausdorff dimension is two[1].

If we look at the series of circles that gradually grows smaller in diameter and extends along the horizontal axis, we come across the following law: the quotients of one diameter and the next tend to a constant with a value of approximately 4.6692… This value, referred to as the Feigenbaum constant, appears in many natural phenomena for reasons that science has, as yet, been unable to explain.

The images are more beautiful and the whole shape becomes clearer when the escape-time algorithm is used to represent the M set using a colour palette. Each colour distinguishes points with a different escape velocity. As an example, a point

---

1 The Hausdorff dimensions of Julia sets vary depending on the value of c. For $c = i$, the Hausdorff dimension is approximately 1.2, and for $c = -0.123 + 0.745i$ it is approximately 1.3934. All of these values are empirically calculated, since the exact dimension for the majority of the sets is unknown.
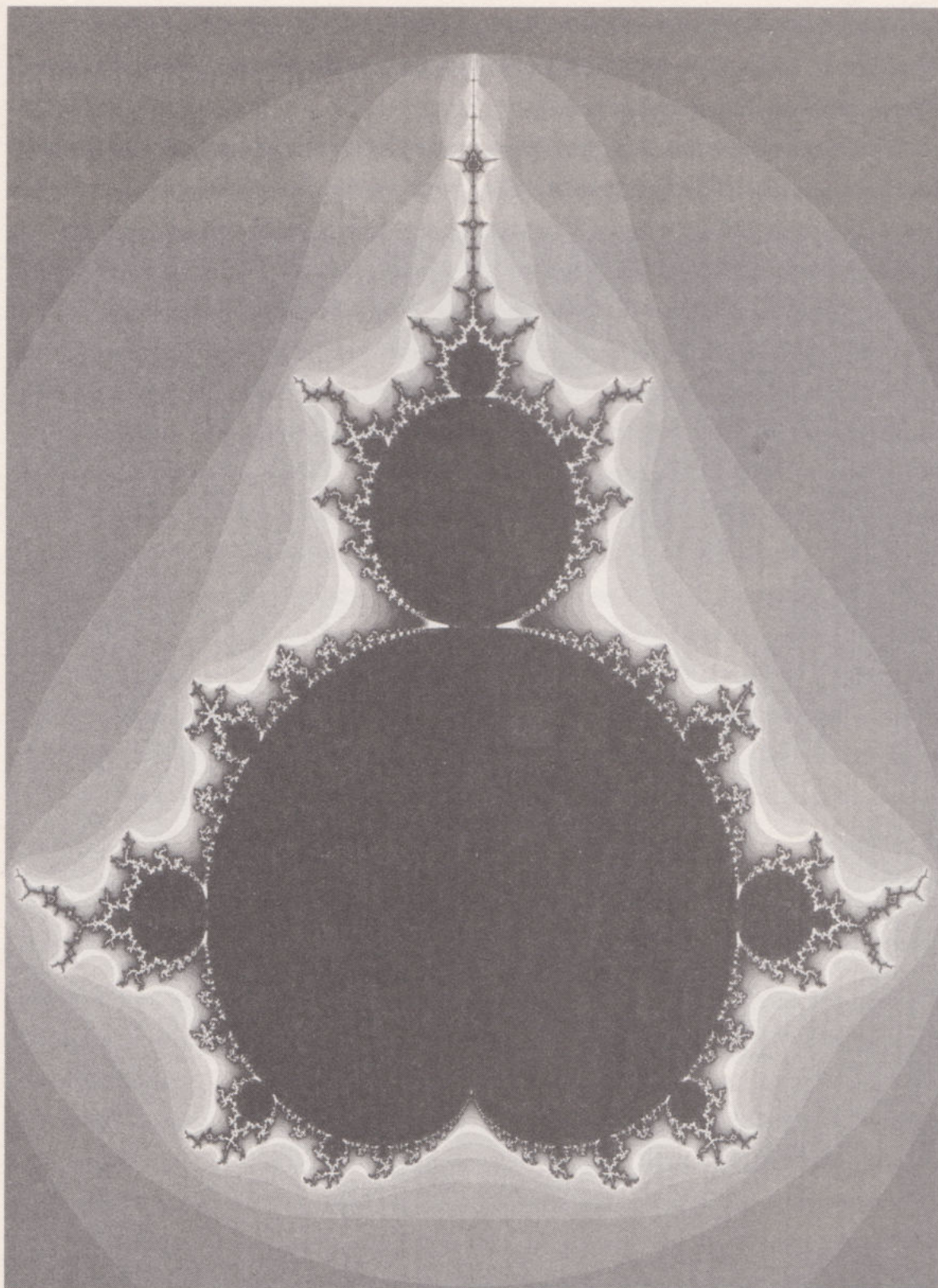
The Mandelbrot set in all its glory. To allow it to be reproduced on the page, it has been displayed vertically instead of horizontally as is more common.

is coloured green if it needs between 11 and 20 iterations to escape from the circle of radius 2, and yellow if it needs between 21 and 30 iterations (see the special colour plates included in this book).

The region known as Seahorse Valley, located between the cardioid and the largest circle, is a sort of marine paradise filled with figures that evoke sea horses linked in thousands of different ways. Throughout the complete length of the plane, we find miniature copies of the larger set which are always wrapped in filaments of varying appearances depending on where we look. The M set appears to be a fractal in the sense that, until now, we have dealt with structures that are repeated at all scales of observation. However the reality is different. With each successive enlargement, the number of structures that are repeated have more and more filaments, allowing us to know which scale we are viewing at. There are serious doubts regarding the self-similarity of the Mandelbrot set. As such, while given any two enlargements of a Julia set, it is impossible to identify at which scale of the plane they have been obtained, the same is not true for the M set. For this reason, the Mandelbrot set is considered to be almost self-similar.

In the 1990s and 2000s, the Chinese mathematician Tan Lei conducted a series of systematic studies of the M set and its complex dynamics, some of which appear in his book *Theme and Variations* (2001). This last image is extremely useful when it comes to tackling the importance of the fractal world beyond the strict limits of mathematics.

Variations on a theme

What is the relationship between the position of a value of c within the Mandelbrot set and the Julia set that is generated using c? The exact answer is unknown. However, it has been possible to empirically observe that the M set contains all the information about the geometry of each Julia set in a reduced and deformed manner, and that, as such, is much more than a simple tool for the classification of connected and disconnected shapes. Thus, for example, all values of c that fall within the main cardioid give rise to a Julia shape with the appearance of a wrinkled circle; if the value of c falls within one of the tangent circles, the Julia set appears lobulated; if it moves towards the filaments, it becomes thinner until it is dendritic and if it crosses the border, it separates into an infinity of segments.

Studying the properties of the M set in greater depth, it is clear that within a given tangent circle, the number of lobules of the associated Julia sets is always

constant. Thus, assigning these numbers to each lobule and studying the lines that result, it is possible to construct a map of the Mandelbrot set.
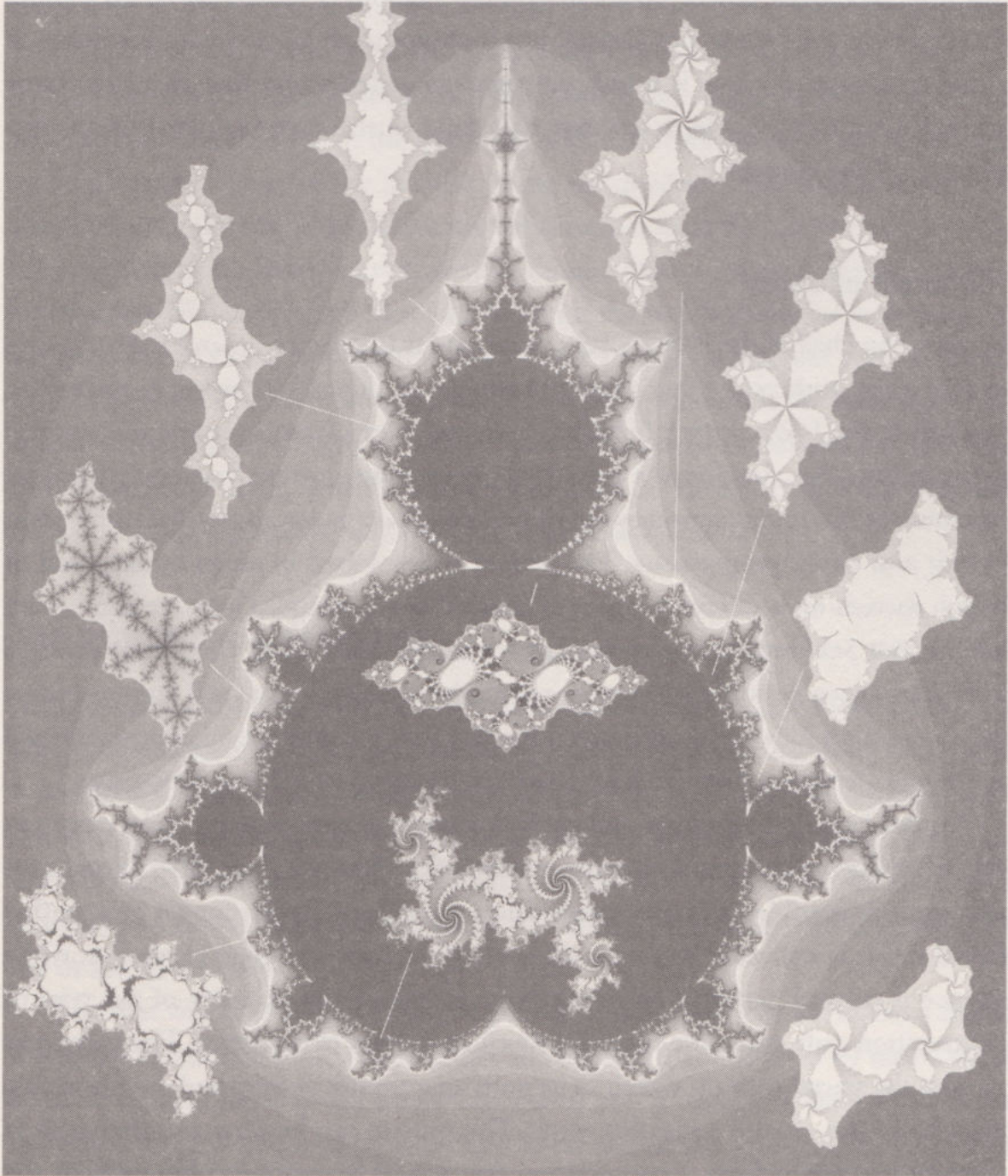
Mandelbrot set with various Julia sets marking the value of c which generates them Once again a vertical orientation has been used to make it easier to appreciate the details.

The creation of this number universe populated with flowers, snails, dragons and filigrees is not an exclusive property to this particular process of iteration. There are many types of iterations of complex numbers that construct their own fractal world. The examples we have already seen show perfectly how a complex structure does not necessarily imply a complex way of forming it. When we bear in mind that reading a single genetic code can generate millions of individual varied features that make up the human population, it is hard not to think that nature behaves in a similar manner. Perhaps it is not just mere coincidence and that, thanks to these analogies, we may be able to discover that most elusive universal law that governs all scales of the cosmos.

The sound of chaos[2]

In the second half of the 20th century, music and mathematics, art and science, began to reconnect with each other thanks to the use of digital processing programmes. Just before the 1920s, Joseph Schillinger, an emigrant Ukranian music theoretician living in the USA, developed a geometric composition system based on periodical movements. He found various ways of relating this concept to rhythm, tone, scale, chords and chord progressions. The system is developed in seven books, each of which focuses on a different aspect of musical composition. There are those who believe that Schillinger's system anticipated music made using computers prior to their appearance. According to some experts, the musician's work still does not enjoy the recognition it deserves, although it has had a huge influence on composers such as George Gershwin, Glenn Miller (both his students) and Benny Goodman.

Models based on equations and random sequences have been the most common way of producing compositions using mathematics. However since 1970 fractals have also made it possible to create composition algorithms. Many people believe that fractal music is not really art, since art is emotional, intuitive and expressive, while science is rational, provable and descriptive. However, the majority of this type of composition only make use of fractal music as a point of departure. The composer moulds the fractal melody, which in its pure state sounds strange and disconcerting, until he or she obtains a pleasing result. The process can be slow and the result is, according to the authors, a composition that would be unimaginable without the

---

2 The following text is directly inspired by *Música Fractal: El Sonido Del Caos,* (Fractal Music: the Sound of Chaos) by Juan Antonio Pérez Ortiz, member of the Department of Computer Languages and Systems at the University of Alicante in Spain.

help of a computer and which, nonetheless, a computer would never have produced in such an elaborate manner.

There are many programs for the generation of fractal music (MusiNum, LMuse, Gingerbread and The Well Tempered Fractal) which can be used to automatically generate pleasant melodies. Phil Thompson, a programmer and amateur musician from Britain, began composing fractal music as a hobby, and in 1998 he published his first album, *Organized Chaos*. His compositions, which he considers as 'discoveries', are based on the Mandelbrot set. Thompson is the creator of the Gingerbread programme, which works in the following way: a starting point, z , is selected, and a transformation is applied to its orbit (using the quadratic function) that converts the points into musical notes. When the orbit escapes from the circle of radius 2, the melody starts again. Based on this, the programme offers a vast quantity of possibilities that give rise to a wealth of results. Without the requirement for knowledge of mathematics or music, it is possible to compose a range of different types of music such as pop, classical, film soundtracks and background sounds for web pages. The author maintains that there is an infinite quantity of material available for composition and defines fractal music as a special form of composition in which the user does not invent the music they hear, but instead discovers it.

The search for a definition

Is there really a rigorous definition of what a fractal is? In *Fractals: Form, Chance and Dimension* Mandelbrot insists that we only have an empirical definition, since there is no theoretical definition that is wholly satisfactory. Sometimes it is said that fractals are objects with a fractional dimension, but this statement is doubly mistaken, since the dimension can be an irrational number (as in the case of the Sierpinsky triangle) and also a whole number (as in the case of space-filling curves or the border of the M set itself).

The best way of defining a fractal is probably by means of its attributes: a fractal is roughly self-similar (the parts look like the whole), is constructed by means of an iterative process, depends on initial conditions and is complex, even if it can be described using a simple algorithm. This is the definition provided by the Briton Kenneth Falconer in his book *Fractal Geometry: Mathematical Foundations and Applications* (1990), in which a fractal structure is defined as one that satisfies one of the following properties:

1. It is too irregular to be described using traditional geometric forms.
2. It has detail at any scale of observation.
3. It is self-similar in some way (exact, approximately or statistically).
4. Its Hausdorff-Besicovitch is strictly greater than its topological dimension.
5. It is defined by means of a simple recursive algorithm.

In 1975, Mandelbrot stated that fractals are shapes that are normally generated by means of repetitive mathematical processes and characterised by not being differentiable since they have a similar appearance at any scale, and have a fractal dimension.

Not everyone was convinced by this formulation and in 1982, Mandelbrot defined a fractal as a set with a Hausdorff dimension that is strictly greater than its topological one. However he himself recognised that this last consideration was not sufficiently general and excluded certain objects that were in fact fractals, such as space-filling curves like the Peano and Hilbert curves that we examined in detail in the first part of the second chapter.[3]

It is also possible to use Barnsley's definition of fractals as IFS attractors, or we can even opt for Judith Cederberg's definition in her *A Course in Modern Geometries* (2001): a set of points that are self-similar in direction, or deterministic or stochastic (depending on chance). Either of these two definitions excludes the M set, something that may or may not be desired. Aware of this Cederberg, writes:

"In the Mandelbrot set, nature (or its mathematics?) gives us a powerful visual analogy of the musical idea, variation on a theme: the same patterns repeated on a given site although the repetition is slightly different… It gives us an excuse to remain interested, because new things appear all the time and nothing is lost because the familiar returns over and over again. As a result of this constant innovation, the Mandelbrot set can be referred to as a *limit* fractal, one that contains many others. Compared with common fractals, it has more structures, its harmonies are richer and the unexpected is even more unexpected."

---

3 The topological dimension of the curves that cover the plane is 2, it is the dimension of the limit of the iteration that is measured.

SELF-DEFINED

There are different ways of classifying fractals in line with the properties they describe. Depending on their degree of self-similarity, for example, fractals can be classified into five broad categories:

- **Exactly self-similar:** This is the most restrictive type since it requires the appearance of the fractal to be identical at different scales. We find it in the Cantor set, the Sierpinsky triangle, the Peano curve, the Koch snowflake, the dragon curve, the Menger, sponge, etc.
- **Linear:** Fractals that are constructed based on affine transformations. This type of fractal contains smaller copies which have been transformed using linear functions, such as the leaf of the Barnsley fern.
- **Strictly self-similar:** This type of fractal contains smaller copies that have been transformed using non-linear functions, such as Julia sets.
- **Quasi self-similar:** This requires fractals to appear to be more or less identical at different levels. This type of fractal contains smaller, distorted copies of itself. Fractals defined by recurrence relations, such as the Mandelbrot set, or the Lyapunov fractal, are normally of this sort.
- **Statistically self-similar:** This type is less strict than self-similarity and requires the fractal to have numerical or statistical numbers which are preserved with the change of scale. Random fractals are one example of this type, and include Brownian motion, Levy's flight, fractal landscapes and Brownian trees.

Nature is not fractal

In the literature about fractals, it is common to find statements such as "nature is fractal". However, the truth is that this is not completely correct. When we say that a border, a tree or the venous network are fractals, what we are really saying is that there are fractal models that can approximate them to a high degree of accuracy. There are no fractals in the real world, just as there are no straight lines or perfect circles.

However, due to the fact that they approximate reality, mathematical models help us to understand them better. In the same way that the theory of relativity

provides a better approximation of the orbit of Mercury than Newtonian mechanics, the fractal model provides a better approximation of the shapes of certain objects than Euclidean geometry, and perhaps also provides a better approximation to the dynamics of real processes.

The Mandelbrot set has an infinite number of details and it would be possible to spend a whole lifetime exploring it, enlargement after enlargement. We can do the same with the real world, starting with molecules and enlarging our field of view to atoms, and from these to neutrons and other subatomic particles. Will we reach the end one day? Or, just like the Mandelbrot set, is there no limit?

Awakening from the deterministic dream

According to the dictionary, chaos is the "utter confusion and disorder", which is assumed prior to the ordering of the cosmos. However scientists are of a different opinion and not all that is said is as pejorative as this definition might suggest.

The mathematical theory of chaos is a branch of an exact science. It does not allow imprecision or lack of definition. While is true is that the name of the theory was inspired by the normal meaning of the word, mathematical chaos is not a wolf, but rather a sheep in wolf's clothing. Fractals allow us to enter into a primordial ocean of structures and systems which, with time, we will come to master.

Fractal geometry and chaos are two interconnected mathematical specialisms and it is difficult to understand one without the other. Fractal geometry studies self-similar and paradoxical patterns and chaos theory is described as the study of the behaviour of unpredictable systems, which it analyses in the search for regularities. Both branches of mathematics, developed in the last 30 years, are related: There are fractals in chaos and chaos can be defined using fractals. But just what is the connection? The general framework from which chaos theory emerges is referred to as dynamical systems theory. A dynamical system consists of two parts: a state (normally expressed using coordinates) and a dynamic (the evolution of the state over the course of time). The evolution of a dynamical system can be displayed in a coordinate space, the positions and speeds of which are known as the phase space. If its evolution is determined by a law or laws (even though its nature is unknown), these are the same at all times and it is possible to deduce a subsequent state based on the previous one. Such a system is referred to as a deterministic dynamical system. The term deterministic means that it is possible to make predictions regarding the future evolution of the system by looking at its past.

One of the most surprising results from physics in recent years is the affirmation that for many deterministic dynamical systems, detailed prediction is impossible over large time intervals since the degree of error increases considerably with each iteration. Such deterministic dynamical systems, which are highly sensitive to relatively small variations, are referred to as chaotic. This extreme sensitivity means that two possible paths that are extremely close at a starting point can diverge considerably over the course of time. It is already known that such randomness appears in systems with a large number of variables, but the fact that it also arises in simple-looking systems was both an unexpected and noteworthy result.

In 1776, the French mathematician Pierre-Simon Laplace categorically confirmed that if the speed and position of all the particles of the Universe were known at a given moment, it would be possible to perfectly establish their past and predict their future. For more than 100 years, this statement appeared to be correct and as a result, it was concluded that, given that everything was determined, at least in terms of power, there could be no free will. In science, this manner of understanding the world, subsequently referred to as Laplace determinism, states that if the laws that govern a certain phenomenon are known, together with the initial conditions, and we have the means to calculate the solution, the it is possible to predict the future of the system with 100% accuracy.

At the end of the 19th century, Poincaré introduced a new perspective when he asked if the Solar System would be stable forever. The French mathematician was the first to suggest the possibility that the behaviour of a system depended appreciably on its initial conditions:

"The behaviour of a system can be analysed, repeating the experiment with the same initial values and under the same conditions such that it is possible to obtain the same results. This leads to the principle of causality. If the same causes produce the same effects, we speak of a strong causality. However, in the majority of cases, it is only possible to achieve initial conditions which are similar, and there is no strong causality. Similar causes produce similar effects."

In 1903, Poincaré made the following statement regarding randomness: "chance is but the expression of man's ignorance."

Twentieth century science has been witness to the collapse of Laplace's determinism for two extremely different reasons. The first is derived from quantum mechanics and Heisenberg's Uncertainty Principle, according to which there is a fundamental limit to the accuracy with which it is possible to measure the position and speed of a particle. However, the reason for large scale unpredictability must be sought somewhere else.

Some large scale phenomena are predictable and others are not. The path of a ball is predictable; that of a balloon carried by the wind is not. Both the ball and the balloon conform to Newton's laws. Why then is the motion of the balloon harder to predict than that of the ball? The motion of water in a calm river is stable, regular and can easily be analysed using equations. However under other conditions, the motion of the water can be turbulent, changeable and unstable. What is the cause of this essential difference? The Soviet physician Lev Landau suggested that as the movement of the water grows faster, the sum of all the oscillations, which may be simple on their own, means that the flow is impossible to predict as a whole. It has been proven, however, that Landau's theory is incorrect. What makes some natural phenomena unpredictable is the growth of errors in measurement and thus the uncertainty regarding their evolution due to chaotic dynamics. This chaotic dynamic is the second reason for the collapse of Laplace determinism. Quantum mechanics confirms that initial measurements can never be completely accurate and chaos claims that imprecisions will dash any ability to make predictions.

A butterfly in Brazil and a moth in Singapore

As we have already noted, Poincaré was the first mathematician to 'discover', or at the very least infer, chaos. In 1890, while studying the astronomical problem of three bodies (consisting of determining, at any instant, the positions and speeds of three bodies with an arbitrary mass that interacted gravitationally, such as the Earth, the Moon and the Sun), Poincaré discovered that it was possible for orbits to exist that were not periodical, nor always growing, nor tending to a fixed point. Eight years later, his compatriot Jacques Hadamard published an influential study on the chaotic motion of three particles in which he showed that their respective paths were unstable and diverged from each other. Hadamard imagined the three particles sliding with friction on a surface with negative curvature, subsequently named Hadamard's billiards

In 1963, the American mathematician and meteorologist Edward Lorenz studied certain equations which he claimed would help him to predict the weather, and attempted to represent their behaviour graphically with the help of computers. Although at the time the machines were considered to be fast, they were extremely slow when compared with those at present, and as such Lorenz went to make himself a cup of tea as they carried out the calculations. Upon his return, he found an extraordinary shape which has been given the name the Lorenz attractor.
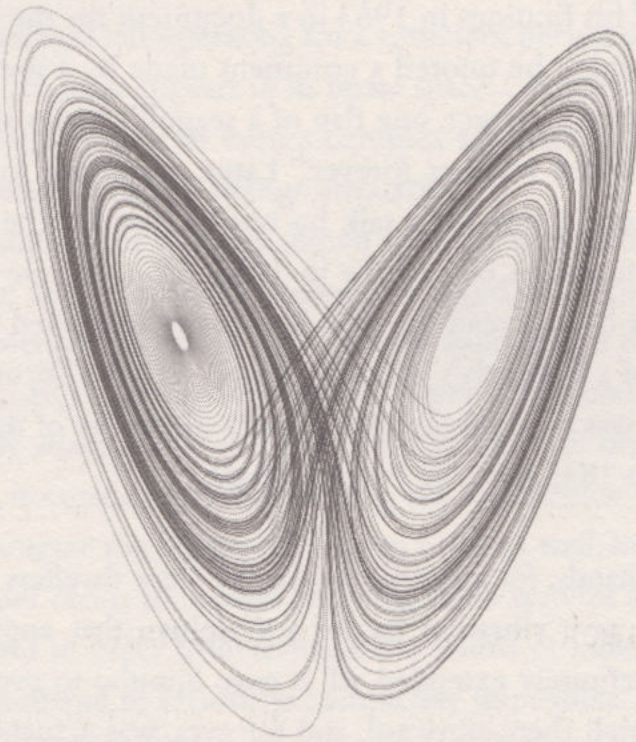
A graphical representation of a Lorenz attractor.

Lorenz thought he had made a mistake when running the program and he repeated it a number of times. After always getting the same result, he began to think that the problem lay in the system itself. After carefully studying it and carrying out tests with different parameters, he observed that, in spite of starting with conditions that were highly similar, the simulations became extremely different. Lorenz believed that the system which he had devised made it possible to set initial conditions to a precision of three decimal places. However, the reality was that the programme worked with six and that the last three decimals were generated

randomly. The American concluded that these small, even minuscule, errors in the starting value multiplied exponentially. Unfortunately, Lorenz published his discoveries in specialist meteorological journals and they went unnoticed for almost a decade.

This verification of what is now referred to as the sensitivity to starting conditions was already present in the work of Hadamard, and was also implicit in Poincaré's studies. However we can go back even further to the work of the Scottish physicist, James Clerk Maxwell, who in 1876 studied different types of chaotic phenomenon, such as the spark that causes a forest fire or the rock that creates an avalanche.[4]

Lorenz published his findings in 1963 in a document for the New York Academy of Sciences in which he quoted a comment made by a meteorological colleague: "If the theory was correct, one flap of a seagull's wings would be enough to alter the course of the weather forever." Later, according to Lorenz himself, finding himself without a title for a talk he was to give for the American Association for the Advancement of Science in 1972, his colleague Felipe Merilees, almost certainly after having read the first document, suggested the following title: "Does the Flap of a Butterfly's Wings in Brazil set off a Tornado in Texas?"

Regardless, there was no doubting that Lorenz was aware of the following fragment of Poe, from his 1845 tale *The Power of Words*:

"We moved our hands, for example, when we were dwellers on the earth, and, in so doing, gave vibration to the atmosphere that engirdled it. This vibration was indefinitely extended, till it gave impulse to every particle of the earth's air, which thenceforward, and for ever, was actuated by the one movement of the hand. This fact the mathematicians of our globe well knew. They made the special effects, indeed, wrought in the fluid by special impulses, the subject of exact calculation — so that it became easy to determine in what precise period an impulse of given extent would engirdle the orb, and impress (for ever) every atom of the atmosphere circumambient."

Lorenz's discovery, later made popular with the name of the Butterfly Effect, suggested the theoretical possibility that the most minuscule of changes in the ini-

---

4 The expression 'dependence on initial conditions' appears to have already been in wide use by 1906 thanks to a publication by the scientist P Dunhem.

tial conditions of a system, such as the movement of the air caused by a butterfly flapping its wings, with respect to the global climate, could set off a chain of events that culminated in significant changes to the system[5].

In spite of the fact that the expression "a butterfly flaps its wings" has survived to the present day, the location of the butterfly, the consequences of its flight and their position has generated a wide debate, the details of which are beyond the scope of this book.

The attraction of chaos

If a dynamical system is left operating for a sufficient period of time, a set of points referred to as the attractor appears in its phase space. Geometrically, an attractor can be a point, a curve, a surface or even a complex set, with an irregular structure, in which case it is referred to as a strange attractor.

The fractal property of chaos is shown through these strange attractors. If the orbits of a strange attractor are successively enlarged, the same self-similarity from fractals can be observed, in which the structure appears and reappears.

On certain occasions, dynamical systems depend on a certain parameter, which makes them easily adaptable to real systems. The value of this parameter is highly important when it comes to understanding the appearance of chaos. For certain values of the parameter, the dynamical system behaves well, however occasionally a small change causes chaos to appear. Studying these systems and the parameter on which they depend in order to determine the points of inflection where a system goes from being ordered to chaotic is completely impossible.

The existence of these families of dynamical systems in which chaos and order coexist forces us to accept that they are a mutually related duality: chaos is always explicitly or implicitly present in any ordered system. Likewise, order is always explicitly or implicitly present in any chaotic system. Even if a system has degenerated into chaos or, on the other hand, it has become ordered and stable, there is always the potential for it to be inverted again.

---

5 The image of the possible global consequences of the innocent flight of a butterfly had already appeared in a short story by Ray Bradbury in 1952 about time travel, and previously in two novels by Charles Hoy Fort in 1923, in which the author speculates that the migration of birds in New York could be responsible for a storm in China.

One example of how order is implicit in chaos is the John Russell soliton. If we throw a stone into a tank of water, a disturbance is generated, causing small ripples to occur, although these will quickly dissipate. In 1834, the Scot John Scott Russell (1808–1882) observed a highly unusual phenomenon. Under certain circumstances, waves joined together to form a new wave with its own characteristics. This new wave, referred to as a soliton, travels hundreds of kilometres without losing its shape. In fact, Russell followed the propagation of this wave along a canal for a number of kilometres, verifying that the current surged without weakening. The Russell soliton is a physical phenomenon in which dispersion and non-linearity interact, causing order in the form of a localised wave.

For practical purposes, the Russell soliton is used to improve the performance of transmissions in optical telecommunications networks. In 1988 solitons were transmitted over more than 4,000 km.

The transition from one laminar state to another turbulent one in a fluid is one of the most representative and widely studied examples of routes to chaos. The Taylor-Couette experiment consists of arranging a fluid between two concentric

The photograph shows the recreation of the Scott Russell soliton that took place on 12 July 1995 at an aqueduct on the Union Canal, close to Heriot-Watt University.

cylinders which rotate at different speeds. When the speed of the rotation of the inside cylinder increases, the fluid ceases to be uniform and breaks up into layers of vortices as if they were doughnuts. The change is even more pronounced when the outer cylinder is rotated in the opposite direction to that of the inner one. Under these circumstances, instead of the vortices, a flow of helicoid spirals appears, such as those which can still be observed in the poles outside barber shops. Changing both speeds opens up a Pandora's box from which a series of wavy and turbulent spiral lines escape. The result varies depending on whether the inner or outer cylinder is accelerated first.

Fractals and chaos are relatively new branches of mathematics, and it would not have been possible to explore them without the power of modern computers. There is no doubt that they have already increased the precision of the description or the classification of 'random' events. However the revolutionary and surprising discovery that certain extremely simple deterministic systems can create randomness suggests an apparent paradox: chaos is deterministic. Generated by fixed rules that do not include any random element, randomness nonetheless occurs.

The discovery of the ubiquity of chaos can be considered as the third great scientific breakthrough of 20th century physics, together with relativity and quantum mechanics.

Bibliography

BARNSLEY, M., *Fractals Everywhere,* San Diego, Academic Press, 1988.

CEDERBERG, J.N., *A Course in Modern Geometries,* New York, Springer, 2000.

PEITGEN, H-O., JURGENS, H. y SAUPE, D., *Chaos and Fractals, New Frontiers of Science,* New York, Springer-Verlag, 1992.

WAHL, B., *Exploring Fractals on the Macintosh,* Reading (Massachusetts), Addison-Wesley, 1995.

Index

A New Look at the World

Fractal geometry

Many natural phenomena exhibit irregular and even chaotic shapes that cannot be analysed using traditional geometry: the sponginess of the clouds, the branches of trees, the zig-zags of lightning bolts. We can seek insights into these apparently random elements of nature in the revolutionary mathematical concept of fractals, a new way of looking at the world based on the maxim, "the whole contains the part and the part the whole."